# Achieved Works

## Paper

Benameur, S. Meraghni, D. Benkhaled, A. Chebana, F and Necir, A. (2016). Complete Flood Frequency Analysis in Abiod Watershed Biskra (Algeria). Natural Hazards, Springer. vol 86, issue 2, page 519-534.

## Presentations at Conferences

• Benameur, S. Meraghni, D. Benkhaled, A. Chebana, F and Necir, A. (November 22, 2011). Frequency Analysis of Extreme Events Flows in the South-East of Algeria (Biskra). International Meeting on Probability and Statistics JIPS. University of Science and Technology Houari Boumediene (USTHB), Algiers. Algeria.

• Benameur, S. Meraghni, D. Benkhaled, A. Chebana, F and Necir, A. (October 18, 2012). Flood Frequency Analysis in Abiod Basin Biskra (Algeria).Séminaire Semestriel en Statistique. Mohamed Khider University, Biskra. Algeria.

• Benameur, S Brahimi, B. Meraghni, D, Necir, A. (March 4, 2013). A Kaplan-Meier Integral for Heavy-tailed Distributions. SéminaireTrimestriel de Statistique (STS). Mohamed Khider University, Biskra. Algeria.

• Benameur, S. Meraghni, D. Benkhaled, A. Chebana, F and Necir, A. (May 6-7, 2014). POT Modelling of Hydrological Extremes in Abiod Basin Biskra (Algeria). Journées de Statistique (JS). Mohamed Khider University, Biskra. Algeria.

*To my grandfather and the memory of my grandmother,*
*To my parents,*
*To my husband and my daughter 'JANA',*
*To my brothers and their kids*
*To all my family,*
*To my husband's family*
*To all those who are dear to me.*

# Acknowledgments

I glorify Allah the Almighty for giving me courage and patience that allowed me to do this work.

I would like to express my greatest gratitude to the people who have helped and supported me throughout my doctorate.

First and foremost I am deeply grateful to my supervisor Prof. Meraghni Djamel, for his valuable support. Without his constructive comments, patience and knowledge this thesis would have not been possible. I honestly could have not wished for a better supervisor.

It has been an extreme (!) pleasure to work with Prof. Abdelkader Benkhaled. Thank you for your passion for the good work, your investment and our discussions about hydrology or otherwise!.

I am very grateful to Prof. Fateh Chebana, for his precious advises, detailed comments and encouragement throughout the years. Thank you for introducing me to the topic of hydrological risk.

Besides my advisor, I would like to thank the examination committee members: Prof. Abdelhakim Necir, Prof. Abdallah Benaissa, Prof. Zaher Mohdeb and the Prof Amar Youkana for their interest in my work and reviewing this thesis, for their insightful comments and encouragement, but also for the hard questions which incited me to widen my research from various perspectives.

A special thank goes to my husband, Benalia Khaled, for always being there for me. Finally, I wish to thank my parents, my sister Nassima and my brothers: Tarek, Chamseddine and Okba, for all the emotional support, understanding, caring they provided, for their never-ending confidence. This thesis is for you.

# Résumé

De nos jours, la gestion du risque occupe une place de plus en plus importante dans le monde socio-économique (commerce, industrie, agriculture, finance, assurance, sociologie, médecine, politique, sport, etc...). D'où la nécessité de se doter de moyens permettant de contrôler un risque donné. On définit alors des quantités théoriques qu'on appelle mesures de risque et qu'on doit être en mesure d'estimer convenablement. Il est évident que pour faire une estimation précise, il faut trouver le modèle théorique le plus approprié aux données. Pour cela, on fait appel à la théorie des valeurs extrêmes qui semble être le meilleur outil permettant la modélisation des événements rares qui influencent grandement les comportements des compagnies pour faire face aux risques dangereux encourus. Le problème est donc d'estimer les différents paramètres d'un modèle de valeurs extrêmes pour pouvoir ensuite aborder l'estimation des mesures de risque.

Ces résultats seront appliqués particulièrement lors des événements hydrologiques extrêmes, tels que les crues et les sécheresses, qui sont l'une des catastrophes naturelles qui se produisent dans plusieurs parties du monde. Ils sont considérés comme étant les risques naturels les plus coûteux en raison des conséquences désastreuses qui se résument essentiellement en pertes en vies humaines et en dégâts matériels. L'objectif principal de la présente étude est d'estimer les événements des crues de Oued Abiod pour des périodes de retour données à la station hydrométrique de M'chouneche située près de Biskra, région semi-aride du Sud-Est de l'Algérie. Cette situation est problématique à plusieurs égards, en raison de l'existence d'un barrage vers l'aval, de la sédimentation et des fuites d'eau à travers le barrage pendant les crues.

Une analyse fréquentielle complète est effectuée sur une série des débits moyens journaliers, par le biais d'outils statistiques classiques ainsi que de techniques récentes. Les résultats obtenus montrent que la distribution de Pareto Généralisée (GPD), pour laquelle les paramètres ont été estimés par la méthode du maximum de vraisemblance (ML), décrit mieux la série analysée. Cette étude indique également aux décideurs l'importance de continuer à surveiller les données à cette station.

**Mots clés :** Valeurs extrêmes; Débits de crues; Analyse fréquentielle; Distribution de Pareto généralisée; Distributions à queues lourdes; Quantiles extrêmes; Evènements rares; Niveau de retour; Mesures de risque; Indice de queue

# Abstract

Nowadays, risk management plays a key role especially in socio-economic world such as: commerce, industry, agriculture, finance, insurance, sociology, medicine, politics and sport, etc. Hence we need some tools in order to control that risk. So we define theoretical quantities that we call risk measures and we will be able to estimate it appropriately. It is obvious that in order to make a precise estimate, we must find the theoretical model most appropriate to the data. This is done using extreme value theory, which seems to be the best tool for modeling rare events that greatly influence the behavior of companies to deal with dangerous risks. This study aims to estimate the various parameters of a model of extreme values in order to be able to approach the estimation of the risk measures.

Those results will be applied especially in extreme hydrological events such as floods, which are one of the natural disasters that occur in several parts of the world. They are regarded as being the most costly natural risks in terms of the disastrous consequences in human lives and in property damages. The main objective of the present study is to estimate flood events of Abiod wadi at given return periods at the gauge station of M'chouneche, located closely to the city of Biskra in a semiarid region of southern east of Algeria. This is a problematic issue in several ways, because of the existence of a dam to the downstream, including the field of the sedimentation and the water leaks through the dam during floods.

A complete frequency analysis is performed on a series of observed daily average discharges, including classical statistical tools as well as recent techniques. The obtained results show that the generalized Pareto distribution (GPD), for which the parameters were estimated by the maximum likelihood (ML) method, describes the analyzed series better. This study also indicates to the decision-makers the importance to continue monitoring data at this station.

**Keywords :** Extreme values; Flood discharges; Frequency analysis; Generalized Pareto distribution; Heavy-tailed distributions; High quantiles; Rare events; Return levels; Risk measures; Tail index.

# Contents

Contents—*Continued*

# List of Figures

# List of Tables

# Introduction

The study of floods is a subject which arouses more and more interest in the field of water sciences. In spite of their low rainfall, the basins of the arid and semiarid areas represent a hydroclimatic context where the overland flow phenomena are significant and feed a network of very active wadis. The activity of these wadis is far from being negligible from the flood in terms of their frequency and intensity. One observes on these rivers exceptional flows, which sometimes, surprise by their magnitude [40]. The Abiod wadi, in the area of Biskra, is a very representative river of these basins. Moreover, the existence of Foum El Gherza dam to the downstream for the irrigation of the palm plantations makes the area more sensitive with regard to the floods. The flood events of the years 1963, 1966, 1971, 1976 and 1989 remain engraved in the memory of the inhabitants. The flood event of September 11–12, 2009, was one of the historic floods in the Zibans area [11]. It rains 80 $mm$ in $24h$, while the annual total of Biskra City reaches 100 $mm$. The damage was 9790 palm trees, 164 flooded houses, 744 destroyed greenhouses, 200 hectares of lost cultures. The last flooding at the time of this drafting paper is that produced in October 29, 2011. All the populations living downstream of the Foum El Gherza dam were evacuated. The floods mainly occur in September and October and especially originate from exceptional storm events.

Describing and studying these situations could help in preventing or at least reducing severe human and material losses. The strategy of prevention of flood risk should be founded on various actions such as risk quantification. On this aspect, various methodological approaches can contribute to this strategy, among which flood frequency analysis (FA). Frequency analysis of extreme hydrological events, such as floods and droughts, is one of the privileged tools by hydrologists for the estimation of such extreme events and their return periods. The main objective of FA approach is the estimation of the probability of exceedance $P(X \geq x_T)$, called hydrological risk, of an event $x_T$ corresponding to a return period $T$ [26]. This process is accomplished by fitting a probability distribution $F$ to large observations in a data set. Two approaches were developed in the context of extreme value theory (EVT). The first one, usually based on the generalized extreme value distribution (GEV), describes the limiting distribution of a suitably normalized annual maximum (AM), and the second uses the generalized Pareto distribution (GPD) to approximate the distribution of peaks over threshold (POT). For more details regarding this theory and its applications, the reader is referred to textbooks such as Embrechts et al. (1997) [47], Reiss and Thomas (1997) [103], Beirlant et al. (2004) [8] and de Hann and Ferriera (2006) [57].

Many FA models should be tested to determine the best fit probability distribution that describes the hydrological data at hand. Specific distributions are recommended in some countries, such as the Log-normal (LN) distribution in China (Bobée 1999) [15]. In the USA, the Log-Pearson type 3 distribution (LP3) has been, since 1967 (National Research Council (NRC) 1988) [91], the official model to which data from all catchments are fitted for planning and insurance purposes. By contrast, the UK endorsed the GEV distribution (Natural Environment Research Council 1975, 1999) [92, 93] up until 1999. The official distribution in this country is now the generalized logistic (GL), as for precipitation in the USA (Willeke et al. 1995) [120]. There are several examples where a number of alternative models have been evaluated for a particular country, for example Kenya (Mutua 1994) [90], Bangladesh (Karim and Chowdhury 1995) [77], Turkey (Bayazıt et al. 1997) [7] and Australia (Vogel et al. 1993) [116]. Nine distributions were used with data from 45 unregulated streams in Turkey by Haktanir (1992) [59] who concluded that two-parameter Log-normal (LN2) and Gumbel distributions were superior to other distributions. Recent research was conducted by Ellouze and Abida (2008) [46] in ten regions of Tunisia. They found that the GEV and GL models provided better estimates of floods than any of the conventional regression methods, generally used for Tunisian floods. Rasmussen et al. (1994) [101] reveal that the POT procedure is more advantageous than the AM in the case of short records. Lang et al. (1999) [84] develop a set of comprehensive practice-oriented guidelines for the use of the POT approach. Tanaka and Takara (2002) [113] has examined several indices to investigate how to determine the number of upper extremes rainfall best for the POT approach.

In the Algerian hydrological context, during the last two decades many authors have used several approaches to study the associated risks. Recently, Hebal and Remini (Hosking 1990) [66] studied flood data from 53 gauge stations in northern Algeria, between 1966 and 2008. They found that 50, 25 and 22% of the samples follow, respectively, the Gamma, Weibull and Halphen A distributions. Bouanani (2005) [17] performed a regional flood FA in the Tafna catchments and concluded that the AM flows fit better to asymmetric distributions such as LP3, Pearson 3 and Gamma. The FA was also used in the sediment context by Benkhaled et al. (2014) [12] where the LN2 distribution was selected in the case of the same station considered in the present study, i.e., M'chouneche gauge station on Abiod wadi.

To our best knowledge, apart from Benkhaled et al. (2014) [12], the flood FA approach has not yet been performed on data collected at this station. The primary aim of this thesis is to perform a FA to the Abiod wadi flow data by the POT approach, based on GPD approximation (Hosking and Wallis 1987) [67]. In methodological terms, all the steps constituting FA are performed from data examination to risk assessment including hypotheses testing and model selection. Due to the high importance of the latter and its impacts, more recent techniques are employed to select the appropriate distribution that fits better to the tail. A relatively large number of known distributions fit well the center of the data, whereas the focus in FA is on the distribution tail. To this end, tail classification and specific graphical tools are employed; see El Adlouni et al. (2008) [45] for more technical details.

This thesis, which focuses on statistical aspects of one-dimensional EVT and its applications in the fields of hydrology, is organized as follows :

**Part I : Preliminary Theory**

**Chapter 1 : Extreme Values.** In this chapter, we provide an overview of the essential definitions and results of EVT. We start by the asymptotic properties of the sum of independent and identically distributed random variables, order statistics and distributions of upper order statistics. Afterwards, we will be interested in the result, first discovered by Fisher and Tippett and later proved in complete generality by Gnedenko; on the fluctuations and asymptotic behavior of the maximum $X_{n,n}$ of a series of independent and identically distributed random variables. A reminder on GEV and GPD approximations, domains of attraction and regular variation functions is given as well.

**Chapter 2 : Tail Index and High Quantile Estimation.** In this chapter, we review existing approaches and methods for the estimation of the Extreme Value Index (EVI) : Parametric approach and semi-parametric approach. We also present in this chapter the different methods and algorithms for the determination of extreme order statistics as well as the estimation of extreme quantiles. In the last section, we discuses risk measurement which is a great part of an organization's overall risk management strategy. Risk measurement is a tool to used to assess the probability of a bad event happening. It can be done by businesses as part of disaster recovery planning and as part of the software development lifecycle. The analysis usually involves assessing the expected impact of a bad event such as a hurricane or tornado. Furthermore, risk analysis also involves an assessment of the likelihood of that bad event occurring.

**Part II : Main Results**

**Chapter 4. Complete Flood Frequency Analysis in Abiod Watershed Biskra (Algeria).** This chapter is designed to estimate flood events of Abiod wadi at given return periods at the gauge station of M'chouneche, located closely to the city of Biskra in a semiarid region of southern east of Algeria. The study area and the data set are briefly described in section 1. Section 2 is devoted to the FA methodology. The results of the study are presented and discussed in the third section.

# Part I

# Preliminary Theory

# Chapter 1

# Extreme Values

## Contents

When we are interested in information about the extreme tail of a distribution, classical statistical tools can not be applied. To this end, extreme values methods were constructed. In this chapter, we will present in a very classical way the extreme value theory, which is the counterpart of the Central Limit Theorem (CLT) for sums. However, while the CLT is concerned with fluctuations around the mean resulting from an aggregation process, the EVT provides results on the asymptotic behavior of the extreme realizations (maxima and minima). Indeed, our starting point will be the order statistics, they are an essential tool in the theory of extreme values. A good reference for the theory and applications of extreme values is the book of Embrechts et al. [47].

## 1.1 Basic Concepts

**Definition 1.1** (Distribution and survival functions).
*If $X$ is a random variable (rv) defined on a probability space $(\Omega, \mathcal{F}, P)$ then, the distribution and survival functions $F$ and $\bar{F}$ are respectively defined on $\mathbb{R}$ by*

$$F(x) := P(X \leq x), \tag{1.1}$$

*and*

$$\bar{F}(x) := 1 - F(x). \tag{1.2}$$

*$\bar{F}$ is also called tail of distribution .*

**Definition 1.2** (Sum and arithmetic mean).
*Let $X_1, X_2, \cdots$ be a sequence of random variables (rv's) that are independent and identically distributed (iid) defined on the same probability space. For any integer $n \geq 1$, we define the sum and the corresponding arithmetic mean respectively by*

$$S_n := \sum_{i=1}^{n} X_i, \tag{1.3}$$

*and*

$$\bar{X}_n := S_n / n. \tag{1.4}$$

### 1.1.1 Law of Large Numbers

The laws of large numbers indicate that as the number of randomly-drawn observations $n$ in a sample increases, the statistical characteristics of the draw (the sample) closer to the statistical characteristics of the population. They are of two types; Weak laws involving convergence in probability and strong laws relating to almost safe convergence.

**Theorem 1.1** (Law of large numbers).
*Let* $(X_1, X_2, \cdots, X_n)$ *be a sample of a rv* $X$*, with finite expected value* $(E|X| < \infty)$*, then*

$$\text{Weak Law : } \bar{X}_n \xrightarrow{p} E(X) \text{ as } n \longrightarrow \infty,$$
$$\text{Strong Law : } \bar{X}_n \xrightarrow{a.s.} E(X) \text{ as } n \longrightarrow \infty.$$

**Definition 1.3** (Empirical distribution function).
*The empirical distribution function of a sample* $(X_1, X_2, \cdots, X_n)$ *is defined by*

$$F_n(x) := \frac{1}{n}\sum_{i=1}^{n} \mathbb{I}_{\{X_i \leq x\}}, \ x \in \mathbb{R}, \tag{1.5}$$

*where* $\mathbb{I}_A$ *denotes the indicator function of the set* $A$*.*

The application of the strong law of large numbers on $F_n(x)$ gives the following result.

**Corollary 1.1.**

$$F_n(x) \xrightarrow{a.s.} F(x) \text{ as } n \longrightarrow \infty, \text{ for every } x \in \mathbb{R}.$$

The result of this corollary can be strengthened in the following fundamental result in nonparametric statistics, known under the name of theorem Glivenko-Cantelli.

**Theorem 1.2** (Glivenko-Cantelli).
*The convergence of* $F_n$ *to* $F$ *is almost surely uniform, i.e.*

$$\sup_{x\in\mathbb{R}} |F_n(x) - F(x)| \xrightarrow{a.s.} 0 \text{ as } n \longrightarrow \infty.$$

The proof of Theorems 1.1 and 1.2 could be found in any standard textbook of probability theory such as [9, chapter 4, page 268]

### 1.1.2 Central Limit Theorem

The CLT states that a sum of $n$ rv's independently drawn from a common distribution function $F(x)$ with finite variance, converge to the normal distribution as $n$ goes to infinity.

**Theorem 1.3** (CLT).
*If $X_1, X_2, \cdots$ is a sequence of rv's iid of mean $\mu$ and finite variance $\sigma^2$, then*

$$(S_n - n\mu)/\sigma\sqrt{n} \xrightarrow{d} \mathcal{N}(0,1) \quad \text{as } n \longrightarrow \infty.$$

The proof of this Theorem could be found in any standard book of statistics (see e.g., [106, page 66]).

## 1.2 Order Statistics

The extreme value theory is directly linked to that of order statistics. This section gathers definitions and the results that we need throughout this thesis. For more details, we refer to books ([2], [32] and [39]).

**Definition 1.4** (Order Statistics).
*If the random variables $X_1, X_2, \cdots, X_n$ are arranged in increasing order of magnitude and then written as*

$$X_{1,n} \leq X_{2,n} \leq \cdots \leq X_{n,n},$$

*the random variable $X_{i,n}$ is called the $i$th order statistics ($i = 1, \cdots, n$).*

In the following we will assume that $X_i$ are independent and identically distributed random variables from a continuous population with cumulative distribution function (cdf) $F$ and probability density function (pdf) $f$.

### 1.2.1 Distribution of An Order Statistics

The distribution function of the $k$th order statistics $X_{k,n}$, for $1 \leq k \leq n$, denoted by $F_{X_{k,n}}$ is obtained as follows

$$\begin{aligned}
F_{X_{k,n}}(x) &= P(X_{k,n} \leq x) \\
&= P(\text{at least } k \text{ observations among } X_i \text{ are } \leq x) \\
&= P\left(\sum_{i=1}^{n} \mathbb{1}_{\{X_i \leq x\}} \geq k\right) \\
&= \sum_{i=k}^{n} \binom{n}{i} F^i(x)\left[1-F(x)\right]^{n-i}, \ -\infty < x < +\infty, \qquad (1.6)
\end{aligned}$$

since $\sum_{i=1}^{n} \mathbb{1}_{\{X_i \leq x\}}$ follows the binomial distribution ; $Bin(n, F(x))$.

The density function is then

$$f_{X_{k,n}}(x) = \frac{1}{B(k, n-k+1)} F^{k-1}(x)\left\{1-F(x)\right\}^{n-k} f(x), \qquad (1.7)$$

where

$$B(k, n-k+1) = \frac{n!}{(k-1)!\,(n-k)!}.$$

Particular cases of interest in the extreme value theory are

- The maximum, $X_{n,n}$, with distribution and density functions respectively

$$\begin{aligned}
F_{X_{n,n}}(x) &= P(X_{n,n} \leq x) = P\left\{X_1 \leq x, \cdots, X_n \leq x\right\} \\
&= \prod_{i=1}^{n} P(X_1 \leq x) \\
&= \left\{F(x)\right\}^n, -\infty < x < +\infty, \qquad (1.8)
\end{aligned}$$

  and

$$f_{X_{n,n}}(x) = nF^{n-1}(x)f(x). \qquad (1.9)$$

- The minimum, $X_{1,n}$, with distribution and density functions respectively

$$\begin{aligned}
F_{X_{1,n}}(x) &= P(X_{1,n} \leq x) = 1 - P(X_{1,n} > x) \\
&= 1 - P\left\{X_1 > x, \cdots, X_n > x\right\} \\
&= 1 - \left\{\bar{F}(x)\right\}^n, -\infty < x < +\infty, \qquad (1.10)
\end{aligned}$$

  and

$$f_{X_{1,n}}(x) = n\left(1-F(x)\right)^{n-1} f(x). \qquad (1.11)$$

- Let $U_{1,n}, \cdots, U_{n,n}$ be the order statistics corresponding to $n$ iid rv's $U_1, \cdots, U_n$ from a uniform distribution in the unit interval ($\mathcal{U}(0,1)$ distribution), then the $i$th order statistics, $U_{i,n}$, follows a beta distribution with parameters $i$ and $n-i+1$, i.e., $U_{i,n} \sim Be(i, n-i+1)$

$$f_{U_{i,n}}(x) = \frac{1}{B(i, n-i+1)} x^{i-1} (1-x)^{n-i}, x \in (0,1). \tag{1.12}$$

As the uniform distribution on the unit interval is symmetric with respect to $1/2$, (that is $U \overset{d}{=} 1-U$)

$$U_{i,n} \overset{d}{=} 1 - U_{n-i+1,n}.$$

The proof of those densities results could be found in the textbook [2, pages 10,12-14], or see e.g, [47, pages 183-184].

### 1.2.2 Joint Density of Two Order Statistics

It can be checked that the joint density of two order statistics $(X_{j,n}, X_{k,n})$ with $(1 \leq j < k \leq n)$ is

$$\begin{aligned} f_{X_{j,n},X_{k,n}}(x,y) &= \frac{n!}{(j-1)!\,(k-j-1)!\,(n-k)!} \\ &\quad .F^{j-1}(x)\left\{F(y) - F(x)\right\}^{k-j-1}\left\{1 - F(y)\right\}^{n-k} f(x) f(y), \\ &\qquad -\infty < x < y < +\infty. \end{aligned} \tag{1.13}$$

In particular the joint density of the maximum and the minimum $(X_{1,n}, X_{n,n})$ is

$$f_{X_{1,n},X_{n,n}}(x,y) = n(n-1)\left\{F(y) - F(x)\right\}^{n-2} f(x) f(y), -\infty < x < y < +\infty. \tag{1.14}$$

### 1.2.3 Joint Density of All the Order Statistics

The joint density of all the order statistics is

$$f_{X_{1,n},X_{2,n},\ldots,X_{n,n}}(x_1, x_2, \ldots, x_n) = n! \prod_{i=1}^{n} f(x_i), \quad -\infty < x_1 < \cdots < x_n < +\infty. \tag{1.15}$$

From this joint density we could have obtained, the density of a single order statistics, or the joint density of two order statistics.

The detailed proof of the above joint densities could be found in the book of Davis et Nagaraga [32].

### 1.2.4 Some Properties of Order Statistics

First, we introduce the quantile function (or generalized inverse).

**Definition 1.5** (Quantile function).
*Let $F$ be a distribution function. The quantile function is*

$$Q(s) = F^{\leftarrow}(s) := \inf\{x \in \mathbb{R} \;:\; F(x) \geq s\}, \; 0 < s < 1. \tag{1.16}$$

For any cdf $F$, the quantile function is non-decreasing and right-continuous. If $F$ is continuous, then $Q$ is continuous. If $F$ is strictly increasing, then $Q$ is the inverse function $F^{-1}$. The most important property of the quantile function is :

**Theorem 1.4** (Quantile transformation).
*Let $X$ be a rv with cdf $F$. Let $U \sim \mathcal{U}(0,1)$. Then, the cdf of the rv $Q(s)$ is $F$, or in other words*

$$X \stackrel{d}{=} Q(U). \tag{1.17}$$

**Proposition 1.1** (Quantile transformation).

- *Let $X_{1,n}, X_{2,n}, \cdots, X_{n,n}$ be the order statistics of $n$ iid observation from a rv $X$ with distribution function $F$. Consider the transformed rv $Y = g(X)$, with $g$ a Borel measurable function. As the order is preserved by non-decreasing function, we have*

$$(Y_{1,n}, Y_{2,n}, \cdots, Y_{n,n}) \stackrel{d}{=} (g(X_{1,n}), g(X_{2,n}), \cdots, g(X_{n,n})), \tag{1.18}$$

  *for any non-decreasing function $g$.*

- *In particular, let $(U_1, \cdots, U_n)$ be a sample from the standard uniform rv and $(U_{1,n}, \cdots, U_{n,n})$ the corresponding ordered sample*

$$(X_{1,n}, \cdots, X_{n,n}) \stackrel{d}{=} (Q(U_{1,n}), \cdots, Q(U_{n,n})) \tag{1.19}$$

- *From (1.19), we have*

$$X_{i,n} \stackrel{d}{=} Q(U_{i,n}), \; i = 1, \cdots, n.$$

- *When $F$ is continuous, we have*

$$F(X_{i,n}) \stackrel{d}{=} U_{i,n}, \; i = 1, \cdots, n. \tag{1.20}$$

See [102], Theorem 1.2.5, page 17, for the proof.

**Proposition 1.2** (Moments).
*The* $m$*th* $(m = 1, 2, \cdots)$ *moment of the* $i$*th* $(i = 1, \cdots, n)$ *order statistics is*

$$\begin{aligned} \mu_{i,n}^{(m)} &= EX_{i,n}^{m} = \frac{1}{B(i, n-i+1)} \int_{-\infty}^{+\infty} x^m \{F(x)\}^{i-1} \{F(x)\}^{n-i} f(x)\, dx \\ &= \frac{1}{B(i, n-i+1)} \int_{0}^{1} \{Q(s)\}^m s^{i-1} (1-s)^{n-i}\, ds. \end{aligned} \tag{1.21}$$

**Proposition 1.3** (Markov property).
*When the original iid variables* $X_1, X_2, \cdots, X_n$ *are ordered, the corresponding order statistics* $X_{1,n}, X_{2,n}, \cdots, X_{n,n}$ *are no longer independent. When* $F$ *is continuous, the dependence structure can be described by a Markov chain. In other words, we have for* $i = 2, \cdots, n$

$$P(X_{i,n} \le x \,|X_{1,n} = x_1, \cdots, X_{i-1,n} = x_{i-1}) := P(X_{i,n} \le x \,|X_{i-1,n} = x_{i-1}).$$

The proofs of these results are straightforward and could be found in [2, page 14, Theorem 3.4.1 page 48].

### 1.2.5 Properties of Uniform and Exponential Spacings

The three following theorems (of which the proof can be found in the book [39] (chapter 5)), give the properties of uniform and exponential spacings.

Let $U_{1,n}, \cdots, U_{n,n}$ be the order statistics corresponding to $n$ iid rv's $U_1, \cdots, U_n$ from a uniform distribution in the unit interval. The statistics $S_i$ defined by

$$S_i := U_{i,n} - U_{i-1,n} \ \ (i = 1, \cdots, n+1), \tag{1.22}$$

where by convention $U_{0,n} = 0$, $U_{n+1,n} = 1$, are called the uniform spacings for this sample.

**Theorem 1.5.**
$(S_1, \cdots, S_n)$ *is uniformly distributed over the simplex*

$$A_n := \left\{ (x_1, \ldots, x_n) : x_i \ge 0, \sum_{i=1}^{n} x_i \le 1 \right\}. \tag{1.23}$$

**Theorem 1.6** (Pyke, 1965, 1972 [98, 99] )**.**
*Let $E_1, E_2, \cdots, E_{n+1}$ be a sequence of iid exponential rv's, then*

$$\{S_1, \ldots, S_{n+1}\} \stackrel{d}{=} \left\{ \frac{E_1}{\sum_{i=1}^{n+1} E_i}, \frac{E_2}{\sum_{i=1}^{n+1} E_i}, \ldots, \frac{E_{n+1}}{\sum_{i=1}^{n+1} E_i} \right\}. \tag{1.24}$$

**Theorem 1.7** (Sukhatme, 1937 [112])**.**
*Let $E_{1,n}, \cdots, E_{n,n}$ be the order statistics corresponding to a sequence of $n$ iid rv's from a standard exponential distribution $E_1, E_2, \cdots, E_n$. If we define $E_{0,n} = 0$, then the normalized exponential spacings*

$$(n - i + 1)\left(E_{i,n} - E_{i-1,n}\right), 1 \leq i \leq n,$$

*are iid exponential random variables. Also*

$$(E_{1,n}, \cdots, E_{n,n}) \stackrel{d}{=} \left( \frac{E_1}{n}, \frac{E_1}{n} + \frac{E_2}{n-1}, \cdots, \frac{E_1}{n} + \frac{E_2}{n-1} + \cdots + \frac{E_n}{1} \right). \tag{1.25}$$

**Theorem 1.8** (Malmquist, 1950 [87])**.**
*Let $U_{1,n}, \cdots, U_{n,n}$ be the order statistics of $U_1, \cdots, U_n$, a sequence of iid uniform $[0,1]$ random variables. Then, if $U_{n+1,n} = 1$*

$$\left\{ \left( \frac{U_{i,n}}{U_{i+1,n}} \right)^i, \ 1 \leq i \leq n \right\} \stackrel{d}{=} \left\{ U_i, 1 \leq i \leq n \right\}, \tag{1.26}$$

## 1.3 Limit Distributions and Domains of Attraction

When modeling the maxima (or minima) of a random variables, extreme value theory plays the same fundamental role as the Central Limit theorem plays when modeling the sum of random variables. In both cases, the theory tells us what the limiting distributions are. Generally there are two approaches can be considered in identifying extremes in real data.

The first, called block maxima approach, consists of dividing the series into non-overlapping blocks of the same length and choosing the maximum from each block and fitting the GEV to the set of block. The assumption that the extreme observations are iid is viable in this case. But the choice of block size can be critical. The choice amounts to a trade-off between bias and variance : blocks that are too small mean that approximation by the limit model is likely to be poor, leading to bias in estimation and extrapolation ; large blocks generate few block maxima, leading to large estimation variance. Pragmatic considerations often lead to the adoption of blocks of length one year (Annual Maxima).

The second approach consists of choosing a given threshold (high enough) and considering the extreme observations exceeding this threshold. This approach based on the GPD approximation is called the peaks-over-threshold (POT) approach. The choice of the threshold is also subject to a trade-off between variance and bias. By increasing the number of observations for the series of maxima (a lower threshold), some observations from the centre of the distribution are introduced in the series, and the index of tail is more precise (less variance) but biased. On the other hand, choosing a high threshold reduces the bias but makes the estimator more volatile (fewer observations). The problem of dependent observations is also present. Detailed and technical introduction can be found in de Haan and Ferreira (2006) [57], Embrechts et al. (1997) [47] and Coles (2001) [28].

The main analytic tool of EVT is the theory of regularly varying functions. So, before proceeding to the presentation of extreme value theory, we provide an introduction to concepts such as regularly varying functions, among others, which are commonly used in EVT and are necessary for a better comprehension of the logic and of the results of this theory.

### 1.3.1 Regular Variation

The concept of regular variation is widely used in EVT to describe the deviation from pure power laws. Regular variation of the tails of a distribution appears as a condition in various theoretical results of probability theory, so in domain of attraction. In this section, we summarize some of the main results of regular variation theory. An encyclopedic treatment of regular variation can be found in Bingham et al [14].

**Definition 1.6** (Regularly varying functions).
*A positive, Lebesgue measurable function $h$ on $(0,\infty)$ is regularly varying at infinity of index $\rho \in \mathbb{R}$, we write $h \in \mathcal{R}_\rho$, iff*

$$\lim_{t\to\infty} \frac{h(tx)}{h(t)} = x^\rho, \text{ for all } x > 0. \tag{1.27}$$

*When $\rho = 0$, function $h$ is said to be slowly varying at infinity.*

**Proposition 1.4** (Regular and slow variations).
*Any regularly varying function $h$ can be decomposed as*

$$h(x) := x^\rho L(x), \tag{1.28}$$

*with $L$ is called slowly varying function.*

Notice that a slowly varying function is essentially a regularly varying function with index 0. Typical examples are positive constants or functions converging to a positive constant, logarithms and iterated logarithms.

The three foundation stones of the theory of regular variation are the Karamata representation theorem, the uniform convergence theorem and the characterization theorem, which identifies the crucial concept of the index of regular variation.

**Theorem 1.9** (Karamata representation).
*If $h \in \mathcal{R}_\rho$ for some $\rho \in \mathbb{R}$, then*

$$h(x) = c(x) \exp\left\{ \int_A^x \frac{r(t)}{t} dt \right\}, x \geq A \tag{1.29}$$

*for some $A > 0$, where $c$ and $r$ are measurable functions, such that $c(x) \to c_0 \in (0,\infty)$ and $r(x) \to \rho$ as $x \to \infty$. The converse implication also holds.*

**Proof.** See Resnik [104, Corollary 2.1, page 29]. □

**Proposition 1.5.**
*From the representation theorem we may conclude that for regularly varying $h$ with index $\rho \neq 0$, as $x \to \infty$,*

$$h(x) \to \begin{cases} \infty & if \quad \rho > 0, \\ 0 & if \quad \rho < 0. \end{cases}$$

The proof of this proposition is detailed in Resnik [104, Proposition 2.6, page 32].

**Theorem 1.10** (Uniform convergence). *If $h \in \mathcal{R}_\rho$ (in the case $\rho > 0$, assuming $h$ bounded on each interval $(0, x]$, $x > 0$), then for $0 < a \leq b < \infty$ relation (1.27) holds uniformly in $x$*

**(a)** *on each $[a, b]$ if $\rho = 0$,*

**(b)** *on each $(0, b]$ if $\rho > 0$,*

**(c)** *on each $[a, \infty)$ if $\rho < 0$.*

**Proof.** See e.g. Bingham et al [14, Theorem 1.5.2, page 22]. □

The following result of Karamata is also very useful, since it is often used in proofs of theorems of extreme value theory. It says that integrals of regularly varying functions are again regularly varying functions, or more precisely, one can take the slowly varying function out of the integral.

**Theorem 1.11** (Karamata, 1933).
*Let $l$ be a slowly function, bounded in $[x_0, \infty)$ for some $x_0 \geq 0$. Then*

**1)** *for $\rho > -1$*

$$\int_{x_0}^{x} t^\rho l(t)\, dt \sim (\rho + 1)^{-1} x^{\rho+1} L(x), \text{ as } x \to \infty,$$

**2)** *for $\rho < -1$*

$$\int_{x}^{\infty} t^\rho l(t)\, dt \sim -(\rho + 1)^{-1} x^{\rho+1} L(x), \text{ as } x \to \infty.$$

**Corollary 1.2.**
*The conclusions of Karamata's theorem can alternatively be formulated as follows. Suppose $h \in \mathcal{R}_\rho$ and $h$ is locally bounded on $[x_0, \infty)$ for some $x_0 \geq 0$. Then*

**(a)** *for $\rho > -1$*

$$\lim_{x\to\infty} \frac{\int_{x_0}^{x} h(t)\, dt}{x h(x)} = \frac{1}{\rho + 1},$$

**(b)** *for* $\rho < -1$

$$\lim_{x\to\infty} \frac{\int_x^{\infty} h(t)\,dt}{xh(x)} = -\frac{1}{\rho+1}.$$

The following result is crucial for the differentiation of regularly varying functions.

**Theorem 1.12** (Monotone density).
*Let* $K(x) = \int_0^x k(y)\,dy$ *(or* $\int_x^{\infty} k(y)\,dy$*) where* $k$ *is ultimately monotone (i.e.* $u$ *is monotone on* $(A,\infty)$ *for some* $A > 0$*). If*

$$K(x) \sim cx^{\rho} L(x), x \to \infty,$$

*with* $c \geq 0$*,* $\rho \in \mathbb{R}$ *and* $l \in \mathcal{R}_0$ *then*

$$k(x) \sim c\rho x^{\rho-1} L(x), x \to \infty.$$

*For* $c = 0$ *the above relations are interpreted as*

$$K(x) = o(x^{\rho} L(x)) \text{ and } uk(x) = o\left(x^{\rho-1} L(x)\right).$$

**Definition 1.7** (Regularly varying rv and distribution).
*A non-negative rv* $X$ *and its distribution are said to be regularly varying with index* $\rho \geq 0$ *if the right distribution tail* $\bar{F} \in \mathcal{R}_{-\rho}$.

**Proposition 1.6** (Regularly varying distributions).
*Assume that* $F$ *is a continuous cdf (with pdf* $f$*) such that* $F(x) < 1$ *for all* $x \geq 0$.

**(a)** *Suppose for some* $\rho > 0$, $\lim_{x\to\infty} xf(x)/\bar{F}(x) = \rho$*, then* $f \in \mathcal{R}_{-(1+\rho)}$ *and consequently* $\bar{F}(x) \in \mathcal{R}_{-\rho}$.

**(b)** *Suppose* $f \in \mathcal{R}_{-(1+\rho)}$ *for some* $\rho > 0$*, then* $\lim_{x\to\infty} xf(x)/\bar{F}(x) = \rho$*. The latter statement also holds if* $\bar{F} \in \mathcal{R}_{-\rho}$ *for some* $\rho > 0$ *and* $f$ *is ultimately monotone.*

**(c)** *Suppose* $X$ *is a regularly varying non-negative rv with index* $\rho > 0$*. Then*

$$\begin{aligned} EX^p &< \infty \text{ if } p < \rho, \\ EX^p &= \infty \text{ if } p > \rho. \end{aligned}$$

**(d)** *Suppose* $\bar{F} \in \mathcal{R}_{-\rho}$ *for some* $\rho > 0$, $p \geq \rho$

$$\lim_{x \to \infty} \frac{x^p \bar{F}(x)}{\displaystyle\int_0^x y^p dF(y)} = \frac{p-\rho}{\rho}.$$

*The converse also holds in the case that* $p > \rho$. *If* $p = \rho$ *one can only conclude that* $\bar{F}(x) = o(x^{-\rho} L(x))$ *for some* $L \in \mathcal{R}_0$.

Now, we return to the main topic of this chapter, which is the presentation of extreme value theory.

### 1.3.2 GEV Approximation

Result (1.8) is of no immediate interest, since it simply says that for any fixed $x$ for which $F(x) < 1$, we have $P(X_{n,n} \leq x) \to 0$ (see e.g. Coles (2001) [28]). In the EVT we are interested in the limiting distribution of normalized maxima. The mathematical foundation is the class of extreme value limits laws originally derived by Fisher and Tippett (1928) [48] and Gnedenko (1943) [50] and summarized in the following theorem, which plays a key role in EVT.

**Theorem 1.13** ((Fisher and Tippet, 1928) [48], (Gnedenko, 1943) [50])**.** *Let* $X_1, X_2, \ldots, X_n$ *be a sequence of iid rv's. If there exist sequences of constants* $a_n > 0$ *and* $b_n \in \mathbb{R}$ *and some non-degenerate distribution function* $H$ *(i.e., some distribution function which does not put all its mass at a single point), such that*

$$a_n^{-1}(X_{n,n} - b_n) \xrightarrow{d} H, \text{ as } n \to \infty, \tag{1.30}$$

*then* $H$ *belongs to one of the following three standard extreme value distributions*

*Gumbel :* $\Lambda(x) := \exp(-e^{-x}), \ x \in \mathbb{R}.$

*Fréchet :* $\Phi_\alpha(x) := \begin{cases} 0, & x \leq 0 \\ \exp(-x^{-\alpha}), & x > 0 \end{cases}$ *and* $\alpha > 0$.

*Weibull :* $\Psi_\alpha(x) := \begin{cases} \exp(-(-x)^\alpha), & x \leq 0 \\ 1, & x > 0 \end{cases}$ *and* $\alpha > 0$.

Sketches of proofs, extensions, choice of normalizing constants, and applications, can be found in Embrechts et al. (1997) [47, page 122], Kotz and Nadarajah (2000) [80], and Coles (2001) [28].

In the figure below, we give a visual inspection of the form of the limiting df's.

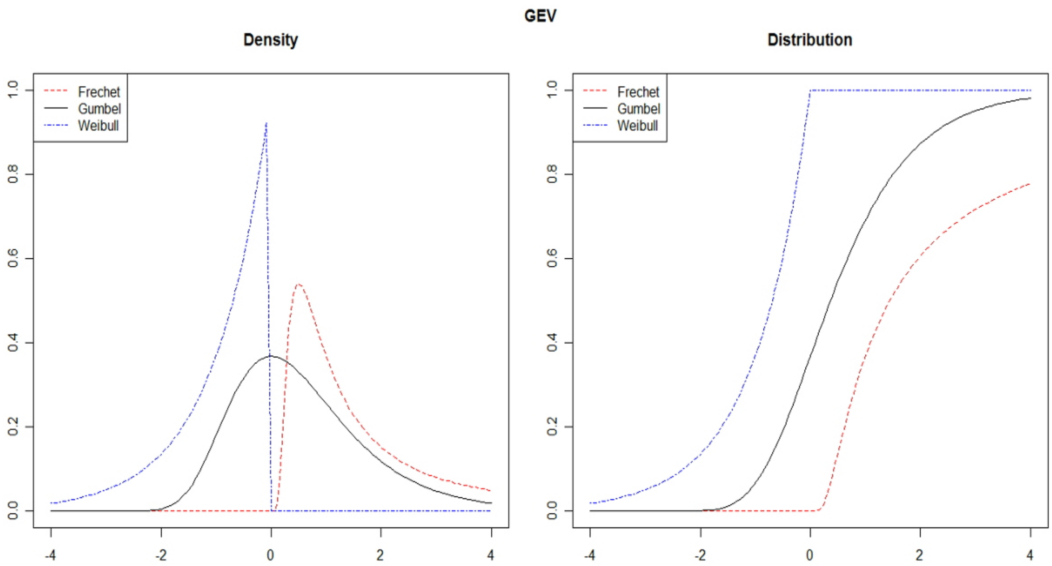



FIGURE 1.1. Density and Distributions of extreme value distributions

In accordance with von Mises (1936) [117] and Jenkinson (1955) [73], we can obtain a one-parameter representation of the three standard distributions. This representation is known as the standard generalized extreme value(GEV) distribution.

**Definition 1.8** (GEV Distribution)**.**
*The standard GEV distribution is given by*

$$H_\gamma(x) := \begin{cases} \exp\left(-(1+\gamma x)^{-1/\gamma}\right) & for \quad \gamma \neq 0,\ 1+\gamma x > 0, \\ \exp(-\exp(-x)) & for \quad \gamma = 0,\ x \in \mathbb{R}. \end{cases} \tag{1.31}$$

*where the parameter $\gamma$ is called "shape parameter", though it is often referred to as "extreme value index" (EVI) or "tail index" of $F$.*

The corresponding pdf $h_\gamma$ is defined by

$$h_\gamma(x) := \begin{cases} H_\gamma(x)(1+\gamma x)^{-1/\gamma-1} & if \quad \gamma \neq 0,\ 1+\gamma x > 0, \\ \exp(-x-\exp(-x)) & if \quad \gamma = 0,\ x \in \mathbb{R}. \end{cases} \tag{1.32}$$

The related location-scale family $H_{\gamma,\mu,\sigma}$ can be introduced by replacing the argument $x$ above by $(x-\mu)/\sigma$ for $\mu \in \mathbb{R},\ \sigma > 0$ ; that is $H_{\gamma,\mu,\sigma}(x) := H_\gamma\left(\dfrac{x-\mu}{\sigma}\right)$.

One can derive the correspondence between the GEV distribution and the three standard extreme value df's. Specifically

$$\begin{aligned}
\Lambda(x) &= H_0(x), & x \in \mathbb{R}. \\
\Phi_\alpha(x) &= H_{1/\alpha}(\alpha(x-1)), & x > 0. \\
\Psi_\alpha(x) &= H_{-1/\alpha}(\alpha(x+1)), & x < 0.
\end{aligned}$$

In other words, a parametric family $\{H_\gamma, \gamma \in \mathbb{R}\}$ is introduced. It provides a unifying representation for the three types of limit distributions.

$$H_\gamma = \begin{cases} \Psi_{-1/\gamma} & if \quad \gamma < 0, \\ \Lambda & if \quad \gamma = 0, \\ \Phi_{1/\gamma} & if \quad \gamma > 0. \end{cases}$$

In order to explore the necessary conditions for the existence of a limiting distribution function $H$, it is useful to adopt a systematic approach towards the set of df.'s whose maxima have the same limiting df. So, we introduce the notion of maximum domains of attraction.

### 1.3.3 Maximum Domains of Attraction

The fact that the extreme value distribution functions are continuous on $\mathbb{R}$, from (1.8), the relation (1.30), for some $\gamma \in \mathbb{R}$, is equivalent to

$$\lim_{n\to\infty} P\left(\frac{X_{n,n} - b_n}{a_n} \leq x\right) = \lim_{n\to\infty} F^n(a_n x + b_n) = H_\gamma(x), x \in \mathbb{R}. \tag{1.33}$$

**Definition 1.9** (Domain of attraction)**.**
*If (1.33) holds for some normalizing constants $a_n > 0$, $b_n \in \mathbb{R}$ and non-degenerate distribution function $H$, we say that the rv $X$ and its distribution function $F$ belong to the domain of attraction of the extreme value distribution $H$, and we write $F \in \mathcal{D}(H)$.*

#### General Characterization

In practice, it is often more convenient, not to work on the cdf $F$ itself, but on the tail function.

**Definition 1.10** (Tail quantile function).
*The tail quantile function is defined by*

$$U(t) := Q(1-1/t) = \left(1/\bar{F}\right)^{\leftarrow}(t), \ 1 < t < \infty, \tag{1.34}$$

*where $Q$ is the quantile function of the cdf $F$ defined in (1.16).*

**Definition 1.11** (Upper endpoint).
*The upper (or right) endpoint of the cdf $F$ is defined as follows*

$$x_F := \sup\{x \in \mathbb{R} : F(x) < 1\} \leq \infty. \tag{1.35}$$

**Proposition 1.7** (Limit of $X_{n,n}$).

$$X_{n,n} \overset{a.s}{\to} x_F \ \text{as } n \to \infty.$$

**Proposition 1.8** (Caracterization of $\mathcal{D}(H)$).
*The df $F \in \mathcal{D}(H)$, with normalizing constants $a_n > 0$ and $b_n \in \mathbb{R}$, iff*

$$n\bar{F}(a_n x + b_n) \to -\log H_\gamma(x) \ \text{ as } n \to \infty. \tag{1.36}$$

*When $H_\gamma(x) = 0$, the limit is interpreted as $\infty$.*

**Proposition 1.9.**
*$F \in \mathcal{D}(H_\gamma)$ iff, for all $x > 0$, with $(1+\gamma x) > 0$,*

$$\lim_{t \to x_F} \frac{\bar{F}(t + xb(t))}{\bar{F}(t)} = \begin{cases} (1+\gamma x)^{-1/\gamma} & if \quad \gamma \neq 0, \\ \exp(-x) & if \quad \gamma = 0, \end{cases} \tag{1.37}$$

*where $b(x)$ is a positive measurable function.*

**Theorem 1.14.**
*The df $F$ with right endpoint $x_F \leq \infty$ belongs to the maximum domain of attraction of $H_\gamma$ ($F \in \mathcal{D}(H_\gamma)$) iff, for $x, y > 0$, and $y \neq 1$*

$$\begin{aligned} \lim_{s \to \infty} \frac{U(sx) - U(s)}{U(sy) - U(s)} &= \lim_{s \to \infty} \frac{U(sx) - U(s)}{a(s)} \frac{a(s)}{U(sy) - U(s)} \\ &= \begin{cases} \dfrac{x^\gamma - 1}{y^\gamma - 1} & if \quad \gamma \neq 0, \\ \dfrac{\log x}{\log y} & if \quad \gamma = 0. \end{cases} \end{aligned} \tag{1.38}$$

**Theorem 1.15** (Caracterization of $\mathcal{D}(\Phi_\gamma)$).
*The df* $F \in \mathcal{D}(\Phi_\gamma)$, $\gamma > 0$, *iff*

$$\bar{F}(x) = x^{-\gamma} L(x), \tag{1.39}$$

*for some slowly varying function* $L$.

Every $F \in \mathcal{D}(\Phi_\gamma)$ has an infinite right endpoint $x_F = +\infty$. Essentially, $\mathcal{D}(\Phi_\gamma)$ embrace all the distribution with right tails regularly varying with index $-\gamma$ (e.g. Pareto, Cauchy, Student and Burr distribution). These df's are called Pareto-type or heavy-tailed distributions. In this case, the normalizing constants can be chosen as $a_n = U(n) = Q(1 - 1/n)$ and $b_n = 0$. Hence

$$a_n^{-1} X_{n,n} \xrightarrow{d} \Phi_\gamma \text{ as } n \to \infty.$$

**Proof.** See Embrechts et al [47, Theorem 3.3.7, page 131]. □

**Theorem 1.16** (Caracterization of $\mathcal{D}(\Psi_\gamma)$).
*The df* $F \in \mathcal{D}(\Psi_\gamma)$, $\gamma < 0$ *iff* $x_F < +\infty$ *and*

$$\bar{F}\left(x_F - x^{-1}\right) = x^{-\gamma} L(x), \tag{1.40}$$

*for some slowly varying function* $L$. *In this case, the normalizing constants can be chosen as* $a_n = x_F - U(n) = x_F - Q(1 - 1/n)$ *and* $b_n = x_F$. *Hence*

$$a_n^{-1}(X_{n,n} - x_F) \xrightarrow{d} \Psi_\gamma \text{ as } n \to \infty.$$

The proof of Theorem 1.16 is similar to that of the preceding theorem, see Embrechts et al [47, Theorem 3.3.12] for the reciprocal one.

The uniform, beta, inverse of Pareto belong to the domain of attraction of Weibull.

The Gumbel's domain of attraction is more difficult to treat, since there is no direct linkage between the tail and the regular variation notion such as the domains of attraction of Fréchet and Weibull. We will find the extensions of the regular variation that take into account a complete characterization of $\mathcal{D}(\Psi_\gamma)$. The Gumbel class contains the exponential, normal, lognormal, gamma and classical Weibull distributions.

**Definition 1.12** (von Mises function).
*The df $F$ is called a von Mises function with auxiliary function $a$ if there exists some $z < x_F$ such that*

$$\bar{F}(x) := c \exp\left\{-\int_z^x \frac{1}{a(t)} dt\right\}, \quad z < x < x_F, \tag{1.41}$$

*where $c > 0$ is some positive constant, and $a$ is a positive absolutely continuous function (with respect to Lebesgue measure) with density $\acute{a}$ satisfying*

$$\lim_{x \to x_F} \acute{a}(x) = 0.$$

As an example of the von Mises function, the exponential distribution function with parameter $\lambda$, $\bar{F}(x) = e^{-\lambda x}$, the auxiliary function is $a(x) = 1/\lambda$.

**Proposition 1.10** (von Mises function's properties).
*Let $F$ be a von Mises function with auxiliary function $a$. Then*

*(a) $F$ is absolutely continuous on $(z, x_F)$ with positive pdf $f$. The auxiliary function can be chosen as $a(x) = \bar{F}(x) / f(x)$.*

*(b) If $x_F = \infty$, then $\bar{F} \in \mathcal{R}_{-\infty}$ and*

$$\lim_{x \to x_F} \frac{x f(x)}{\bar{F}(x)} = \infty. \tag{1.42}$$

*(c) If $x_F < \infty$, then $\bar{F}(x_F - x^{-1}) \in \mathcal{R}_{-\infty}$ and*

$$\lim_{x \to x_F} \frac{(x_F - x) f(x)}{\bar{F}(x)} = \infty. \tag{1.43}$$

**Theorem 1.17** (von Mises Conditions).
*Let $F$ be an absolutely continuous df with density $f$*

*(a) If*

$$\lim_{x \to \infty} \frac{x f(x)}{\bar{F}(x)} = \gamma > 0, \tag{1.44}$$

*then $F \in \mathcal{D}(\Phi_\gamma)$*

*(b) Assume that the density function $f$ is positive on some finite interval $(z, x_F)$, with $x_F < \infty$. If*

$$\lim_{x \to x_F^-} \frac{(x_F - x) f(x)}{\bar{F}(x)} = \gamma > 0, \tag{1.45}$$

*then $F \in \mathcal{D}(\Psi_\gamma)$.*

(c) *Let $F$ be a df with right endpoint $x_F \leq \infty$, such that for $z < x_F$, $F$ has the representation*

$$\bar{F}(x) = c(x) \exp\left\{ -\int_z^x \frac{g(t)}{a(t)} dt \right\}, \ z < x < x_F, \tag{1.46}$$

*where $g$ and $c$ are some positive functions, such that $c(x) \to c > 0$, $g(x) \to 1$ as $x \to x_F$, and $a(x)$ is a positive, absolutely continuous function (with respect to Lebesgue measure) with density $\acute{a}$ having $\lim_{x \to x_F} \acute{a}(x) = 0$, then $F \in \mathcal{D}(\Lambda)$. In this case, we can choose $b_n = Q(1 - 1/n)$ and $a_n = a(b_n)$ as normalizing constants. A possible choice for $a$ is*

$$a(x) = \int_x^{x_F} \frac{\bar{F}(t)}{\bar{F}(x)} dt, \ x < x_F. \tag{1.47}$$

*The function $a(x)$ is usually called mean-excess function, defined below as (1.49).*

The proof of the last result can be found in [104, Proposition 1.4 and Corollory 1.7].

### 1.3.4 GPD Approximation

Let $X_1, X_2, ...$be a sequence of iid rv's, having marginal df $F$. Modeling only block maxima can be a wasteful approach to extreme value analysis if one block happens to contain more extreme events than another. Let $u$ be a real "sufficiently large" and less than the end point ($u < x_F$), called the threshold. It is natural to regard as extreme events those of the $X_i$ that exceed some high threshold $u$. The second result of the EVT, introduced by de Haan (1993) [56], involves estimating the conditional distribution of the excess over a given threshold. The method of excesses is based on an approximation of excesses distribution over the threshold $u$ of the real rv $X$, i.e. the conditional distribution of the positive real random variable $X - u$ given that $X > u$.

**Definition 1.13** (Distribution and mean of excess)**.**
*Let $X$ be a random variable with a distribution function $F$. The distribution of excess over the threshold $(u < x_F)$ is defined as*

$$F_u(y) := P(X - u \leq y \,|X > u) = 1 - \frac{\bar{F}(u+y)}{\bar{F}(u)}, \ 0 < y < x_F - u, \tag{1.48}$$

*and the corresponding mean*

$$e\left(u\right):=E\left(X-u\left|X>u\right.\right),\ u<x_F, \tag{1.49}$$

*which is also expressed in the form*

$$e\left(u\right)=\frac{1}{\bar{F}\left(u\right)}\int_u^{x_F}\bar{F}\left(x\right)dx,\ u<x_F.$$

*Note that the function* $h\left(u\right)=F\left(u\right)e\left(u\right)$ *is differentiable and* $\dfrac{h'\left(u\right)}{h\left(u\right)}=-1/e\left(u\right)$, *from where* $z<x<x_F$

$$\bar{F}\left(x\right)=\frac{e\left(z\right)}{e\left(x\right)}\exp\left\{-\int_z^x\frac{du}{e\left(u\right)}\right\}. \tag{1.50}$$

The necessary and sufficient condition (proposition 1.9), so that $F\in D\left(H_\gamma\right)$, admits a probabilistic interpretation since

$$\frac{\bar{F}\left(t+xb\left(t\right)\right)}{\bar{F}\left(t\right)}=P\left(\frac{X-t}{b\left(t\right)}>x\left|X>t\right.\right)=1-F_u\left(b\left(u\right)x\right), \tag{1.51}$$

for some positive function $b$. Therefore

$$\lim_{t\to x_F}P\left(\frac{X-t}{b\left(t\right)}>x\left|X>t\right.\right)=\begin{cases}\left(1+\gamma x\right)^{-1/\gamma} & if\quad \gamma\neq0,\\ \exp\left(-x\right) & if\quad \gamma=0,\end{cases} \tag{1.52}$$

For any $x$, with $1+\gamma x>0$. This limit motivates the definition of the essential distribution to the modeling of excesses. Once the threshold is estimated, the conditional distribution $F_u$ is approximated by a Generalized Pareto Distribution (GPD).

**Definition 1.14** (Standard GPD)**.**

The standard Generalized Pareto Distribution is defined, for $\gamma\in\mathbb{R}$, by

$$G_\gamma\left(x\right):=\begin{cases}1-\left(1+\gamma x\right)^{-1/\gamma} & if\quad \gamma\neq0,\\ 1-\exp\left(-x\right) & if\quad \gamma=0,\end{cases} \tag{1.53}$$

where $x\in\mathbb{R}_+$ if $\gamma\geq0$, and $x\in[0,-1/\gamma[$ if $\gamma<0$.

The standard GPD can be extended to a more general family, by replacing the argument $x$ by $\left(x-\mu\right)/\sigma$, where $\mu\in\mathbb{R}$ and $\sigma>0$ are respectively the location and scale parameters. The standard GPD corresponds to the case $\mu=0$ and $\sigma=1$. The GPD with null location parameter and arbitrary scale parameter $\sigma>0$ plays an important role in statistical analysis of extreme events, this specific family is defined as follows.

**Definition 1.15** (Generalized Pareto Distribution).
*The distribution function of the GPD is given by*

$$G_{\gamma,\sigma}(x) = \begin{cases} 1 - \left(1 + \dfrac{\gamma x}{\sigma}\right)^{-1/\gamma} & if \quad \gamma \neq 0, \\ 1 - \exp(-x/\sigma) & if \quad \gamma = 0, \end{cases} \tag{1.54}$$

where

$$x \in D(\gamma, \sigma) = \begin{cases} [0, \infty) & if \quad \gamma \geq 0, \\ [0, -\sigma/\gamma] & if \quad \gamma < 0. \end{cases}$$

In the figure below, we give a visual inspection of the form of the GPD for different values of $\gamma$.

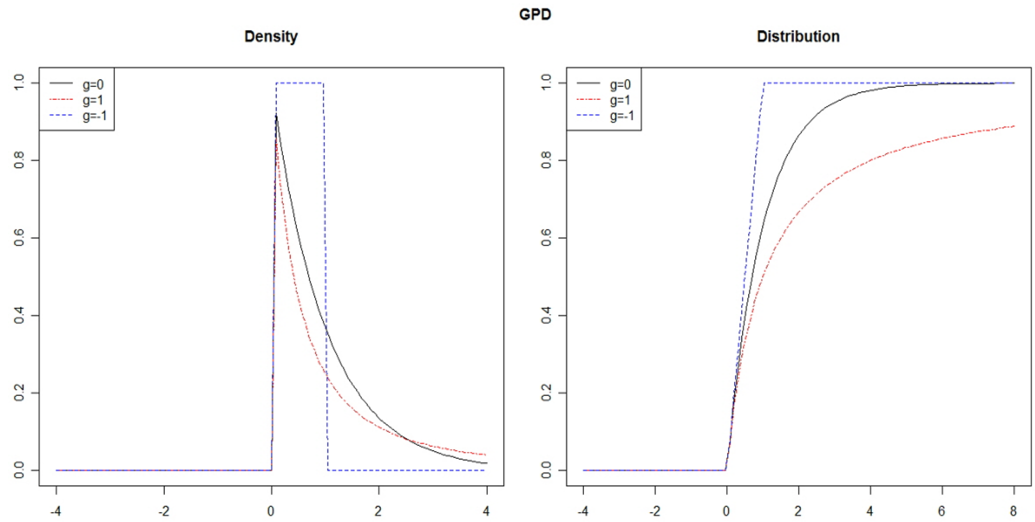



FIGURE 1.2. Density and Distribution of Generalized Pareto Distribution for different values of $\gamma$.

**Theorem 1.18** (GPD Properties).

(a) *Assume that $X$ is a rv having generalized Pareto distribution with parameters $\gamma \in \mathbb{R}$ and $\sigma > 0$. Then $EX < \infty$ iff $\gamma < 1$. In this case, the mean excess function is linear. More precisely, for $u < x_F$*

$$e(u) = \frac{\sigma + \gamma u}{1 - \gamma}, \ \sigma + \gamma u > 0.$$

*If $\gamma < 1/r$ with $r \in \mathbb{N}$, then*

$$EX^r = \frac{\gamma^r \Gamma(\gamma^{-1} - r)}{\gamma^{r+1} \Gamma(1 + \gamma^{-1})} r!,$$

*where* $\Gamma(.)$ *stands for the gamma function,* $\Gamma(t) := \int_0^{+\infty} x^{t-1}e^{-x}dx, t \geq 0.$

(*b*) *Assume that* $N$ *is a rv having Poisson distribution of parameter* $\lambda > 0$ $(N \sim \mathcal{P}(\lambda))$ *independent of an iid sequence* $(X_n)_{n\geq}$ *having a GPD with parameter* $\gamma \in \mathbb{R}$ *and* $\sigma > 0$*. Then*

$$P(M_N \leq x) = \exp\left\{-\lambda\left(1 + \gamma\frac{x}{\sigma}\right)^{-1/\gamma}\right\} = H_{\gamma,\mu,\psi}(x),$$

*where* $M_n = \max(X_1, \cdots, X_N)$*,* $\mu = \sigma\gamma^{-1}(\lambda^\gamma - 1)$ *and* $\psi = \sigma\lambda^\gamma$*.*

For the proof of these properties, one can refer e.g. to the textbook of Embrechts et al [47]

A famous limit result by Pickands [97] and Balkema and de Hann [6], captured in the following theorem, show that the GPD is the natural limiting distribution for excesses over a high threshold.

**Theorem 1.19** ([97], [6]).
*For every* $\gamma \in \mathbb{R}$ *and some positive measurable function* $\sigma(.)$

$$\lim_{u\to x_F} \sup_{0<y<x_F-u} \left|F_u(y) - G_{\gamma,\sigma(u)}(y)\right| = 0. \tag{1.55}$$

*iff* $F \in \mathcal{D}(H_\gamma)$*,* $\gamma \in \mathbb{R}$*.*

Thus, for any distribution $F$ belonging to the maximum domain of attraction of an extreme value distribution, the excess distribution $F_u$ converges uniformly to generalized Pareto distribution as the threshold $u$ is raised.

The proof of the Theorem1.19 must be found in Embrechts et al [47].

**Exemple 1.1** (Standard exponential distribution).
*If* $X_1, X_2, \cdots$ *is a sequence of independent standard exponential variables with distribution* $F(x) = 1 - e^{-x}$ *for* $x > 0$*. Then, by direct calculation*

$$F_u(x) = 1 - \frac{e^{-(u+x)}}{e^{-u}} = 1 - e^{-x} \text{ for } x > 0.$$

*This corresponds to* $\gamma = 0$ *and* $\sigma = 1$ *in (1.54).*

**Exemple 1.2** (Standard Fréchet distribution)**.**

*If $X_1, X_2, \cdots$ is a sequence of independent standard Fréchet variables with distribution $F(x) = \exp(-1/x)$ for $x > 0$. Hence*

$$F_u(x) = \frac{1 - e^{-1/(u+x)}}{1 - e^{-1/u}} \approx 1 - \frac{1/(u+x)}{1/u} = 1 - \left(1 + \frac{x}{u}\right)^{-1} \text{ as } u \to \infty.$$

*This corresponds to $\gamma = 1$ and $\sigma(u) = u$ in (1.54).*

**Exemple 1.3** (Standard uniform distribution)**.**

*If $X_1, X_2, \cdots$ are a sequence of independent uniform $\mathcal{U}(0,1)$ variables with distribution $F(x) = x$ for $0 < x < 1$. Then*

$$F_u(x) = \frac{1 - (u+x)}{1 - u} = 1 - \left(1 + \frac{-x}{1-u}\right).$$

*This corresponds to $\gamma = -1$ and $\sigma(u) = 1 - u$ in (1.54).*

**Chapter 2**

# Estimation of Tail Index, High Quantiles and Risk Measures

## Contents

Estimating parameters constitutes an important task in extreme values theory, since it is a starting point for statistical inference about extreme values of a population. In particular, the extreme value index (EVI) or tail index, measures the right tail's weight of the df $F$, allowing us to describe the behavior of the extreme values of a population. With the estimated EVI, it is possible to estimate other parameters of extreme events like the extreme quantile, the return period and the probability of exceedance of a hight threshold. There are two approaches : a parametric approach and a semi-parametric approach. The parametric approach for modeling extremes is based on the assumption that the data series corresponds to a sample of iid rv's according to one of the extremes distributions. In this case standard estimation methods are applied for the parameters estimation such as; The maximum likelihood (ML) method and the Probability Weighted Moments (PWM) method. In practice, this approach is considered in the case of the AM series (Gumbel 1958) [54]. The other approach, linked to the notion of maximum domain of attractions discussed in section 1.3.3. Indeed, estimation methods based on this approach aim to estimate only the EVI since it is this parameter that determines the shape of the tail distribution. In the literature of EVT there are several semi-parametric techniques for estimating this index. In this chapter we include the Pickands estimator, the Hill estimator, the moment estimator, the kernel-type estimator and the QQ-estimator.

## 2.1 Parameters Estimation Procedures of the GEV Distribution

Again let us consider the GEV distribution of the of maximum whose analytical expression is given by the following system

$$H_\theta(x) := \begin{cases} \exp\left\{-\left(1+\gamma\dfrac{x-\mu}{\sigma}\right)^{-1/\gamma}\right\} & if \quad \gamma \neq 0, 1+\gamma\dfrac{x-\mu}{\sigma} > 0, \\ \exp\left\{-\exp-\left(\dfrac{x-\mu}{\sigma}\right)\right\} & if \quad \gamma = 0, x \in \mathbb{R}, \end{cases} \tag{2.1}$$

where $\theta := (\gamma, \mu, \sigma) \in \ominus \subset \mathbb{R}^2 \times \mathbb{R}_+$.

The probability density function of the GEV is obtained by taking the derivative of the function $H_\theta(x)$ with respect to $x$

$$h_\theta(x) = \begin{cases} \dfrac{1}{\sigma}\left(1+\gamma\left(\dfrac{x-\mu}{\sigma}\right)\right)^{-1/\gamma-1} H_\theta(x) & if \quad \gamma \neq 0,\ 1+\gamma\dfrac{x-\mu}{\sigma} > 0, \\ \dfrac{1}{\sigma}\exp\left[-\dfrac{x-\mu}{\sigma} - \exp\left(-\dfrac{x-\mu}{\sigma}\right)\right] & if \quad \gamma = 0,\ x \in \mathbb{R}. \end{cases} \tag{2.2}$$

### 2.1.1 Parametric Approach

Several estimation methods for the GEV distribution parameters are available in the literature. In this section, we will focus on the most popular ; the Maximum Likelihood (ML) method and the Probability Weighted Moments (PWM).

#### Maximum Likelihood (ML) method

The first method that remains the most popular and which under certain conditions is the most effective is the maximum likelihood (ML) method. It consists in choosing $\theta$ as an estimator of the value that maximizes the likelihood or the log-likelihood function over an appropriate parameter space $\ominus$.

Under the assumption that $(X_1, X_2, \cdots, X_n)$ are independent variables having the GEV distribution, the likelihood function for the GEV parameters is

$$L(\theta; X_1, \cdots, X_n) := \prod_{i=1}^{n} h_\theta(X_i)\, \mathbb{1}_{\{1+\gamma(X_i-\mu)/\sigma>0\}}. \tag{2.3}$$

It is equivalent but often easier to mathematically process the log-likelihood function instead of the likelihood function itself. The log-likelihood function is given by

$$l(\theta; X_1, \cdots, X_n) := \log L(\theta; X_1, \cdots, X_n). \tag{2.4}$$

Therefore

$$\begin{aligned} l(\theta; X_1, \cdots, X_n) &= \sum_{i=1}^{n} \log h_\theta(X_i)\, \mathbb{1}_{\{1+\gamma(X_i-\mu)/\sigma>0\}} \\ &= -n\log\sigma - \left(\frac{1}{\gamma}+1\right)\sum_{i=1}^{n}\log\left(1+\gamma\frac{X_i-\mu}{\sigma}\right) \\ &\quad -\sum_{i=1}^{n}\left(1+\gamma\frac{X_i-\mu}{\sigma}\right)^{-1/\gamma}, \end{aligned} \tag{2.5}$$

which must be maximum, provided that $1+\gamma\left(X_i-\mu\right)/\sigma>0$, for $i=1,\cdots,n$.

The maximum likelihood estimator (ML) is then

$$\hat{\theta}_n=\hat{\theta}_n\left(X_1,\cdots,X_n\right):=\arg\max_{\theta\in\Theta} l\left(\theta;X_1,\cdots,X_n\right). \tag{2.6}$$

In addition, if $l\left(\theta;X_1,\cdots,X_n\right)$ admits partial derivatives with respect to $\gamma,\mu$ and $\sigma$ (resp.), then the ML estimator is often obtained by solving the following equations

$$\frac{\partial l\left(\theta;X_1,\cdots,X_n\right)}{\partial\theta}=0,\ \ \theta=\left(\gamma,\mu,\sigma\right). \tag{2.7}$$

The case where $\gamma=0$ require separate treatment using the Gumbel limit of the GEV distribution. This leads to the log-likelihood function

$$l\left(\theta;X_1,\cdots,X_n\right)=-n\log\sigma-\sum_{i=1}^{n}\exp\left(-\frac{X_i-\mu}{\sigma}\right)-\sum_{i=1}^{n}\frac{X_i-\mu}{\sigma}. \tag{2.8}$$

By differentiating this function relative to the two parameters $\mu$ and $\sigma$ (resp.), we obtain the system of equations to be solved according to

$$\begin{cases} n-\sum_{i=1}^{n}\exp\left(-\frac{X_i-\mu}{\sigma}\right)=0, \\ n+\sum_{i=1}^{n}\frac{X_i-\mu}{\sigma}\left(\exp\left\{-\frac{X_i-\mu}{\sigma}\right\}-1\right)=0. \end{cases} \tag{2.9}$$

However, there is no explicit analytical solution to these nonlinear maximization equations. Thus, for any given dataset numerical procedures and optimization algorithms are used to maximize the likelihood function. Then the calculation of the estimators does not pose serious problems. On the other hand, nothing guarantees their regularities (asymptotically efficient and normal estimators). Smith [109] shows that it is enough that $\gamma>-1/2$ so that the regularity conditions of the ML estimator are fulfilled (for more details, see Castillo, Hadi, Balakrishnan, Sarabia [21]).

### Probability Weighted Moment Method (PWM)

The probability weighted moments (PWM) method is also very popular for fitting the GEV distribution to the dataset. This method is a generalization of the moments method, but with an increasing weight for tail observations [69]. In general, the PWM of a rv $X$ with df $F$, presented by Greenwood, Landwehr, Matalas and Wallis [51], are given by the following quantities

$$M_{p,r,s} := E\left[X^p \left\{F(X)\right\}^r \left\{1 - F(X)\right\}^s\right], \tag{2.10}$$

where $p$, $r$ and $s$ are real numbers. PWM are likely to be most useful when the inverse of the distribution can be written in a closed form, so we can write

$$M_{p,r,s} = \int_0^1 \left\{F^{\leftarrow}\right\}^p F^r \left\{1 - F\right\}^s dF, \tag{2.11}$$

and this is often the most convenient way of evaluating these moments. The specific case of estimation by the PWM method, for the GEV distribution, is studied intensively in Hosking, Wallis and wood [69]. In the case where $\gamma \neq 0$, setting $p = 1$, $r = 0, 1, 2, ...$ and $s = 0$, they would render for the GEV distribution

$$M_{1,r,0} := E\left[X\left\{F(X)\right\}^r\right] = \int_0^1 H_\theta^{\leftarrow}(y)\, y^r dy, \tag{2.12}$$

where $r \in \mathbb{N}$ and for $0 < y < 1$,

$$H_\theta^{\leftarrow}(y) = \begin{cases} \mu - \dfrac{\sigma}{\gamma}\left(1 - (-\log y)^{-\gamma}\right) & if \quad \gamma \neq 0, \\ \mu - \sigma \log(-\log y) & if \quad \gamma = 0. \end{cases} \tag{2.13}$$

Therefore, the PWM for the GEV distribution become

$$M_{1,r,0} = \frac{1}{r+1}\left\{\mu - \frac{\sigma}{\gamma}\left[1 - (r+1)^\gamma \Gamma(1-\gamma)\right]\right\}, \text{ for } \gamma < 1. \tag{2.14}$$

Let $(X_1, X_2, \ldots, X_n)$ be a sample of $n$ iid rv's of GEV, with the associated order statistics $X_{1,n} \leq X_{2,n} \leq \cdots \leq X_{n,n}$. The PWM estimator of $\theta$ is obtained solving the following system of equations resulting from the equation (2.14), with $r = 0, 1, 2$

$$\begin{aligned} M_{1,0,0} &= \mu - \frac{\sigma}{\gamma}\left(1 - \Gamma(1-\gamma)\right), &(2.15)\\ 2M_{1,1,0} - M_{1,0,0} &= \frac{\sigma}{\gamma}\Gamma(1-\gamma)\left(2^\gamma - 1\right), &(2.16)\\ \frac{3M_{1,2,0} - M_{1,0,0}}{2M_{1,1,0} - M_{1,0,0}} &= \frac{3^\gamma - 1}{2^\gamma - 1}. &(2.17) \end{aligned}$$

After replacing $M_{1,r,0}$ by its unbiased estimator (see Landwehr, Matalas and Wallis [83]), given by

$$\hat{M}_{1,r,0} := \frac{1}{n}\sum_{j=1}^{n}\left(\prod_{l=1}^{r}\frac{(j-l)}{(n-l)}\right)X_{j,n}, \tag{2.18}$$

or by the consistent estimator which is asymptotically equivalent

$$\tilde{M}_{1,r,0} := \frac{1}{n}\sum_{j=1}^{n}\left(\frac{j}{n+1}\right)^{r}X_{j,n}. \tag{2.19}$$

we obtain the PWM estimator, $(\hat{\gamma}, \hat{\mu}, \hat{\sigma})$. Note that to obtain $\hat{\gamma}$, the equation (2.17) has to be solved numerically. Then, the equation (2.16) can be solved to obtain $\hat{\sigma}$

$$\hat{\sigma} = \frac{\hat{\gamma}\left(2\hat{M}_{1,1,0} - \hat{M}_{1,0,0}\right)}{\Gamma\left(1-\hat{\gamma}\right)\left(2^{\hat{\gamma}}-1\right)}. \tag{2.20}$$

At the end, given $\hat{\gamma}$, $\hat{\mu}$ can be obtained from equation (2.15)

$$\hat{\mu} = \hat{M}_{1,0,0} + \frac{\hat{\sigma}}{\hat{\gamma}}\left(1 - \Gamma\left(1-\hat{\gamma}\right)\right). \tag{2.21}$$

For more details, see e.g. Beirlant et al [8] and Hosking et al (1985) [69].

### 2.1.2 Semi-Parametric Approach

The semi-parametric approach uses only the characterization of the maximum domain of attraction of the GEV distribution. This approach does not assume the knowledge of the whole distribution but only focus on the distribution tails and the behaviour of extreme values. The case $\gamma > 0$ has got more interest because datasets in most real-life applications, exhibit heavy tails.

In this section, we present some different estimators of the EVI, all based on the order statistics $X_{1,n} \leq X_{2,n} \leq \cdots \leq X_{k,n}$, obtained from the initial series, considering the $k$ highest values, the idea is to have $k \to \infty$ as $n \to \infty$, but without taking too many values of the sample, which leads to impose $k/n \to 0$ as $n \to \infty$.

Incidentally, this implies that the question of the optimal choice of $k$ will arise. Indeed, it is essential to calculate these estimators on the tails of distribution. Choosing a $k$ that is too high creates the risk of taking values that are not extreme, conversely, a sub-sample that is too small does not allow the estimators to reach their level of stability. This sensitive point is discussed in section 2.3 below.

### Pickands' Estimator

The Pickands estimator was introduced in 1975 by Pickands in [97]. It is defined by the statistics

$$\hat{\gamma}^{P}_{k(n)} := \frac{1}{\log 2} \log \left( \frac{X_{n-k(n)+1,n} - X_{n-2k(n)+1,n}}{X_{n-2k(n)+1,n} - X_{n-4k(n)+1,n}} \right). \tag{2.22}$$

We shall give weak consistency and asymptotic properties of $\hat{\gamma}^{P}_{k(n)}$.

**Theorem 2.1** (Weak Consistency of $\hat{\gamma}^{P}_{k(n)}$).
*Let $(X_n)_{n\geq 1}$ be a sequence of iid rv's with df $F \in \mathcal{D}(H_\gamma)$ with $\gamma \in \mathbb{R}$. Then as $k(n) \to \infty$ and $k(n)/n \to 0$*

$$\hat{\gamma}^{P}_{k(n)} \xrightarrow{p} \gamma \text{ as } n \to \infty.$$

**Proof.**
One deduces from the theorem 1.14 (formula (1.38)) that for $\gamma \in \mathbb{R}$, we have with the choice of $t = 2s$, $x = 2$ and $y = 1/2$,

$$\lim_{t\to\infty} \frac{U(t) - U(t/2)}{U(t/2) - U(t/4)} = 2^\gamma.$$

Furthermore, by using the increasing of $U$ which results from the increasing of $F$, one obtains

$$\lim_{t\to\infty} \frac{U(t) - U(tc_1(t))}{U(tc_1(t)) - U(tc_2(t))} = 2^\gamma. \tag{2.23}$$

as soon as $\lim_{t\to\infty} c_1(t) = 1/2$ and $\lim_{t\to\infty} c_2(t) = 1/4$. The basic idea now consists of constructing an empirical estimators for $U(t)$.

To that effect, let $(k(n))_{n\geq 1}$ be a sequence of integers such that $0 \leq k(n) \leq n/4$, $\lim_{n\to\infty} k(n) = \infty$ and $\lim_{n\to\infty} \frac{k(n)}{n} = 0$ (we write $k$ for $k(n)$), let $V_{1,n} \leq \cdots \leq V_{n,n}$ be the order statistics from an iid sample with common standard Pareto df $F_V(x) = 1 - x^{-1}$ for $x \geq 1$.

It is not difficult to see that the sequences

$$\left(\frac{k}{n} V_{n-k+1,n}\right)_{n\geq 1}, \left(\frac{2k}{n} V_{n-2k+1,n}\right)_{n\geq 1} \text{ and } \left(\frac{4k}{n} V_{n-4k+1,n}\right)_{n\geq 1}$$

converge in probability to 1 as $n \to \infty$. In particular, we have the following convergences in probability

$$V_{n-k+1,n} \underset{n\to\infty}{\to} \infty, \ \frac{V_{n-2k+1,n}}{V_{n-k+1,n}} \underset{n\to\infty}{\to} 1/2, \text{ and } \frac{V_{n-4k+1,n}}{V_{n-k+1,n}} \underset{n\to\infty}{\to} 1/4.$$

Combining this with 2.23, it is therefore deduced that the following convergence takes place in probability

$$\frac{U\left(V_{n-k+1,n}\right)-U\left(V_{n-2k+1,n}\right)}{U\left(V_{n-2k+1,n}\right)-U\left(V_{n-4k+1,n}\right)}\underset{n\to\infty}{\to}2^{\gamma}.$$

It remains to determine the distribution of $\left(U\left(V_{1,n}\right),\cdots,U\left(V_{n,n}\right)\right)$. Note that if $x\geq 1$, then $U\left(x\right)=F^{\leftarrow}\left(F_V\left(x\right)\right)$. So, we have

$$\left(U\left(V_{1,n}\right),\cdots,U\left(V_{n,n}\right)\right)=\left(F^{\leftarrow}\left(F_V\left(V_{1,n}\right)\right),\cdots,F^{\leftarrow}\left(F_V\left(V_{n,n}\right)\right)\right),$$

One can deduce from the growth of $F$ that $\left(F_V\left(V_{1,n}\right)\right),\ldots,\left(F_V\left(V_{n,n}\right)\right)$ has the same distribution that the order statistics of $n$ independent uniform rv's over $[0,1]$ . From the proposition 1.1 we can deduce that the random vector

$$F^{\leftarrow}\left(F_V\left(V_{1,n}\right)\right),\ldots,F^{\leftarrow}\left(F_V\left(V_{n,n}\right)\right)$$

has the same distribution that$(X_{1,n},\ldots,X_{n,n})$, the order statistics of a sample of $n$ independent rv's with the df $F$. So the rv

$$\frac{U\left(V_{n-k+1,n}\right)-U\left(V_{n-2k+1,n}\right)}{U\left(V_{n-2k+1,n}\right)-U\left(V_{n-4k+1,n}\right)}\overset{d}{=}\frac{X_{n-k+1,n}-X_{n-2k+1,n}}{X_{n-2k+1,n}-X_{n-4k+1,n}}.$$

Thus this quantity converges in distribution to $2^{\gamma}$ as $n\to\infty$. Since the logarithmic function is continuous on $\mathbb{R}_+^*$, we deduce that the Pickands' estimator converges in distribution to $\gamma$. As $\gamma$ is constant, one also has the convergence in probability. $\square$

**Theorem 2.2** (Asymptotic properties of $\hat{\gamma}^P_{k(n)}$).
*Suppose that $F\in\mathcal{D}\left(H_{\gamma}\right)$, $\gamma\in\mathbb{R}$, $k\to\infty$ and $k/n\to 0$ as $n\to\infty$.*

$(a)$ *Strong consistency : If $k/\log\log n\to\infty$ as $n\to\infty$, then*

$$\hat{\gamma}^P_{k(n)}\overset{a.s.}{\to}\gamma \text{ as } n\to\infty.$$

$(b)$ *Asymptotic normality : Suppose that $U$ has a positive derivatives $U'$and that $\pm t^{1-\gamma}U'\left(t\right)$ (with either choice of sign) is $\Pi$-varying at infinity with auxiliary function $a$.*

*If $k\left(n\right)=o\left(n/g^{\leftarrow}\left(n\right)\right)$ $(n\to\infty)$, where $g\left(t\right):=t^{3-2\gamma}\left(U'\left(t\right)/a\left(t\right)\right)^2$, then*

$$\sqrt{k}\left(\hat{\gamma}^P_{k(n)}-\gamma\right)\overset{d}{\to}\mathcal{N}\left(0,\eta^2\right) \text{ as } n\to\infty,$$

*where*

$$\eta^2 := \left\{ \frac{\gamma^2 \left(2^{2\gamma+1} + 1\right)}{\left\{2\left(2^{\gamma} - 1\right)\log 2\right\}^2} \right. .$$

We refer to Pickands [97] and Dekkers & de Haan (1989) [36] for proofs.

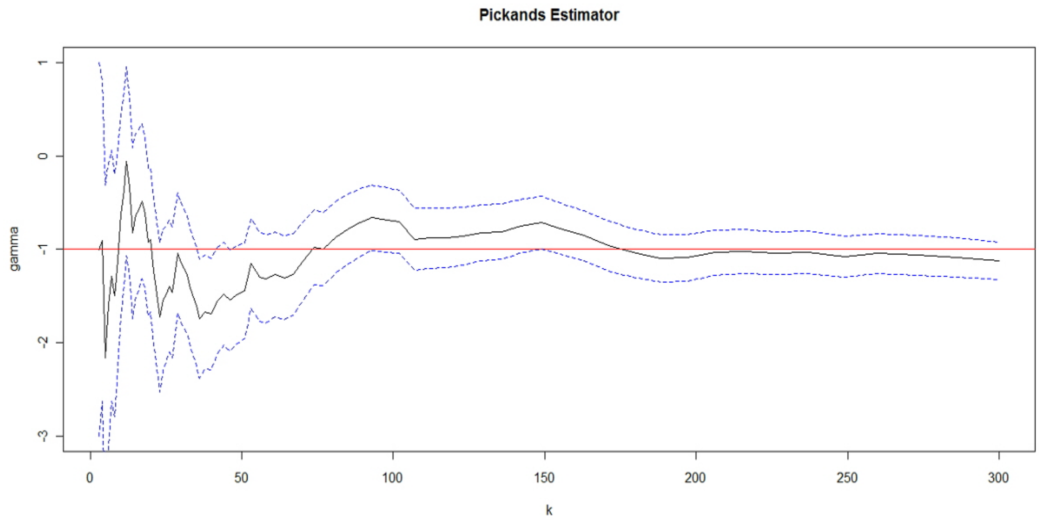



FIGURE 2.1. Pickands estimator, with a confidence interval level of 95%, for the EVI of the standard uniform distribution ($\gamma = -1$) based on 100 samples of 3000 observations.

**Hill's Estimator**

After the Pickands' estimator, Hill (1975) [65] introduced another estimator for $\gamma$, but is restricted to the case of heavy tails df which belong to Fréchet maximum domain of attraction. This most popular estimator is defined for $\gamma > 0$ by the statistics

$$\hat{\gamma}^{H}_{k(n)} := \frac{1}{k(n) - 1} \sum_{i=n-k(n)+2}^{n} \log X_{i,n} - \log X_{n-k(n)+1,n}. \tag{2.24}$$

Or still

$$\hat{\gamma}^{H}_{k(n)} := \frac{1}{k(n)} \sum_{i=1}^{k(n)} \log X_{n-i+1,n} - \log X_{n-k(n),n}. \tag{2.25}$$

Throughout this section, we assume that $\gamma > 0$. In order to construct the Hill estimator, let us start from a preliminary result on the slowly varying functions

**Lemma 2.1.**

*Let $L$ be a slowly varying function. So we have : for all $\rho > 0$, $L(x) = \circ(x^{\rho})$ in $+\infty$ and*

$$\int_x^{\infty} t^{-\rho-1} L(t)\, dt \sim \frac{1}{\rho} x^{-\rho} L(x) \ \text{ in } +\infty.$$

**Proof.**

The proof is based on the representation formula (1.29). Let $\rho > 0$, there exists $x_0$, $M > 0$ such that for $x \geq x_0$, we have $r(x) \leq \rho/2$ and $c(x) \exp \int_z^{x_0} \frac{r(u)}{u} du \leq M$.

One can deduce that for $x \geq x_0$, we have

$$L(x) \leq M \exp \int_{x_0}^{x} \frac{\rho}{2u} du \leq \acute{M} \left(\frac{x}{x_0}\right)^{\rho/2}.$$

We obtain $L(x) = \circ(x^{\rho})$ in $+\infty$.

Let $u \geq 1$. The function $h_x(u) = \left(\frac{L(ux)}{L(x)} - 1\right) u^{-\rho-1}$ is increased in absolute value by $\left(1 + \frac{c(ux)}{c(x)} \exp\left[\int_x^{ux} \frac{r(u)}{u} du\right]\right) u^{-\rho-1}$.

Using the convergences of $c$ and $r$, one deduce that for $x \geq x_0$, the function $|h_x(s)|$ is increased by the function

$$g(u) = u^{-\rho-1}\left(1 + A \exp\left[\int_x^{ux} \frac{r(u)}{u} du\right]\right) \leq \grave{A} u^{-\frac{\rho}{2}-1},$$

where $A$ and $\grave{A}$ are constants that do not depend on $u$. The function $g$ is integrable on $[1, \infty[$, moreover we have $\lim\limits_{x\to\infty} \left(\frac{L(ux)}{L(x)} - 1\right) u^{-\rho-1} = 0$, since $L$ is a slowly varying function . By the dominated convergence theorem, we deduce that

$$\lim_{x\to\infty} \int_1^{\infty} \left(\frac{L(ux)}{L(x)} - 1\right) u^{-\rho-1} du = 0.$$

This implies that $\lim\limits_{x\to\infty} \int_1^{\infty} \frac{L(ux)}{L(x)} u^{-\rho-1} du = \frac{1}{\rho}$ and by change of variable $v = ux$,

$$\lim_{x\to\infty} \frac{1}{x^{-\rho} L(x)} \int_x^{\infty} v^{-\rho-v} L(v)\, dv = \frac{1}{\rho}.$$

One obtain the last property of the lemma. □

**Lemma 2.2.**

*Let* $F \in \mathcal{D}(\Phi_\alpha)$. *We have*

$$\frac{1}{\bar{F}(t)} E\left[(\log X - \log t)\,\mathbb{1}_{(X>t)}\right] \underset{t\to\infty}{\to} \frac{1}{\alpha} = \gamma.$$

**Proof.**

We can deduce from the definition of the slowly varying functions and the theorem 1.15 that $F \in \mathcal{D}(\Phi_\alpha)$, where $\alpha = 1/\gamma$, iff $\lim_{t\to\infty} \frac{\bar{F}(tx)}{\bar{F}(t)} = x^{-\alpha}$ for all $x > 0$. Let us suppose for simplicity that the df of $X$ has the df $f$. By integrating by parts, we have for $t > 1$

$$\begin{aligned} & E\left[(\log X - \log t)\right] \\ = & \int_t^{+\infty} (\log x - \log t)\, f(x)\, dx = \left[-\bar{F}(x)(\log x - \log t)\right]_t^{+\infty} + \int_t^{\infty} \frac{\bar{F}(x)}{x} dx. \end{aligned}$$

In general, the left-hand side and the right one are equal.

The lemma 2.1 gives $\bar{F}(x) = x^{-\alpha} L(x) = \circ\left(x^{-\alpha+\rho}\right)$ with $-\alpha + \rho < 0$. The right-hand side of the equation above reduces to

$$\int_t^{+\infty} \frac{\bar{F}(x)}{x} dx = \int_t^{+\infty} x^{-\alpha-1} L(x)\, dx.$$

We have from the second part of the lemma 2.1

$$\int_t^{+\infty} \frac{\bar{F}(x)}{x} dx = \int_t^{+\infty} x^{-\alpha-1} L(x)\, dx \sim \frac{1}{\alpha} t^{-\alpha} L(t) = \frac{1}{\alpha} \bar{F}(t).$$

Thus, we deduce the lemma. □

It is now necessary to find an estimator of $\bar{F}(t) = E\left[\mathbb{1}_{\{X>t\}}\right]$ and an estimator of $E\left[(\log X - \log t)\,\mathbb{1}_{(X>t)}\right]$.

The strong law of large numbers ensures that $\frac{1}{n}\sum_{i=1}^{n} \mathbb{1}\{X > t\}$ converge a.s. to $\bar{F}(t)$.

It remains to replace $t$ by a quantity which tends to $+\infty$ with $n$. As for the Pickands' estimator, it is natural to replace $t$ by $X_{n-k(n)+1,n}$, where the sequence $(k(n))_{n\geq 1}$ satisfies the following assumptions :

$k(n) \to +\infty$, and $k(n)/n \to 0$ as $n \to \infty$. This last condition ensures that, from the proposition 1.9 and the theorem 1.15, that $X_{n-k(n)+1,n}$ diverges a.s. to infinity.

To lighten the notation, let $k(n) = k$. If we assume that $F$ is continuous, the order statistic is strictly increasing a.s. and we have for $\bar{F}(X_{n-k+1,n})$

$$\bar{F}_n(X_{n-k+1,n}) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}\{X_i > X_{n-k+1,n}\} = \frac{k-1}{n}.$$

The strong law of large numbers ensures that $\frac{1}{n}\sum_{i=1}^{n}(\log X_i - \log t)\,\mathbb{1}_{\{X_i>t\}}$ converge a.s. to $g(t) = E\left[(\log X - \log t)\,\mathbb{1}_{(X>t)}\right]$. Replace again $t$ by $X_{n-k+1,n}$, we get as an estimate of $g(X_{n-k+1,n})$

$$\begin{aligned} & \frac{1}{n}\sum_{i=1}^{n}(\log X_i - \log X_{n-k+1,n})\,\mathbb{1}_{\left\{X_i>X_{n-k+1,n}\right\}} \\ = \; & \frac{1}{n}\left(\sum_{i=n-k+2}^{n}\log X_{i,n} - (k-1)\log X_{n-k+1,n}\right). \end{aligned}$$

Hence, we have

$$\frac{1}{k-1}\sum_{i=n-k+2}^{n}\log X_{i,n} - \log X_{n-k+1,n},$$

which is a good candidate for the estimation of $\gamma$. It is customary to replace $k-1$ with $k$ except in the last term, which does not change the asymptotic result.

Before stating the results on the asymptotic behavior of Hill's estimator, we must impose the second order conditions of regular varying function with a reminder of the first order conditions of regular varying function for heavy-tailed distributions.

**Proposition 2.1** (First order conditions of regular variation ).
*The following assertions are equivalent :*

$(a)$ *$F$ heavy tailed*

$$F \in \mathcal{D}\left(\Phi_{1/\gamma}\right), \gamma > 0. \tag{2.26}$$

$(b)$ *$\bar{F}$ regularly varying at $\infty$ with index $-1/\gamma$*

$$\lim_{t\to\infty}\frac{1-F(tx)}{1-F(t)}=x^{-1/\gamma},\ x>0. \tag{2.27}$$

$(c)$ *$Q(1-s)$ regularly varying at $0$ with index $-\gamma$*

$$\lim_{s\to 0}\frac{Q(1-sx)}{Q(1-s)}=x^{-\gamma},\ x>0. \tag{2.28}$$

$(d)$ *$U$ regularly varying at $\infty$ with index $\gamma$*

$$\lim_{t\to\infty}\frac{U(tx)}{U(t)}=x^{\gamma},\ x>0. \tag{2.29}$$

In a semi-parametric approach, a first order condition is in general not sufficient to study properties of tail parameters' estimators, in particular asymptotic normality. In that case a second order condition is required. The most common one are the following.

**Definition 2.1** (Second order conditions of regular variation)**.**
*The tail of the df $F$, $\bar{F}\in\mathcal{D}(\Phi_\alpha),\alpha=1/\gamma,\gamma>0$, is said to satisfy the second order condition of a regular variation at infinity if one of the following (equivalent) conditions is satisfied :*

$(a)$ *There exist some parameter $\rho\leq 0$, and a function $A^*$ satisfied $\lim_{t\to\infty}A^*(t)=0$ and not changing sign near $\infty$, such that for all $x>0$*

$$\lim_{t\to\infty}\frac{(1-F(tx)/1-F(t))-x^{-\alpha}}{A^*(t)}=x^{-\alpha}\frac{x^{\rho}-1}{\rho}. \tag{2.30}$$

$(b)$ *There exist some parameter$\rho\leq 0$, and a function $A^{**}$ satisfied $\lim_{t\to\infty}A^*(t)=0$ and not changing sign near $0$, such that for all $x>0$*

$$\lim_{s\to 0}\frac{(Q(1-sx)/Q(1-s))-x^{-\gamma}}{A^{**}(t)}=x^{-\gamma}\frac{x^{\rho}-1}{\rho}. \tag{2.31}$$

$(c)$ *There exist some parameter $\rho\leq 0$, and a function $A^*$ satisfied $\lim_{t\to\infty}A^*(t)=0$ and not changing sign near $\infty$, such that for all $x>0$*

$$\lim_{t\to\infty}\frac{(U(tx)/U(t))-x^{\gamma}}{A(t)}=x^{\gamma}\frac{x^{\rho}-1}{\rho} \tag{2.32}$$

if $\rho = 0$, $x^{\rho} - 1/\rho$ is interpreted as $\log x$.

$\rho$ is a second order parameter controlling the speed of convergence of the first order condition.

$A$, $A^{*}$ and $A^{**}$ are regularly varying functions with $A^{*}(t) = A\left(1/\bar{F}(t)\right)$ and $A^{**}(s) = A(1/s)$. Their role is to control the speed of convergence in (2.32), (2.30) and (2.31) (resp.). The relations above may be reformulated respectively

$$\lim_{t\to\infty} \frac{\log(1 - F(tx)) - \log(1 - F(t)) + \alpha \log x}{A^{*}(t)} = \frac{x^{\rho} - 1}{\rho}, \tag{2.33}$$

$$\lim_{s\to 0} \frac{\log Q(1 - sx) - \log Q(1 - s) + \gamma \log x}{A^{**}(t)} = \frac{x^{\rho} - 1}{\rho}, \tag{2.34}$$

and

$$\lim_{t\to\infty} \frac{\log U(tx) - \log U(t) - \gamma \log x}{A(t)} = \frac{x^{\rho} - 1}{\rho}. \tag{2.35}$$

For more details on this issue, we refer to [14] ; [55] ; [49] and [57].

As an example of heavy-tailed distributions satisfying the second-order hypothesis, we have the so called Hall's model.

### Hall's Class of Distribution Functions

A whole class of distribution functions where the index $\gamma$ is of positive, and who is frequently used when one studies the extreme values distributions. This class is given in [61] and it is mentioned by "Hall's model". The df of this class are defined to satisfy

$$\bar{F}(x) = cx^{-1/\gamma}\left(1 + dx^{\rho/\gamma} + o\left(x^{\rho/\gamma}\right)\right) \text{ as } x \to \infty, \tag{2.36}$$

where $\gamma > 0$, $\rho \leq 0$, $c > 0$ and $d \in \mathbb{R} \backslash \{0\}$. Therefore, satisfied quantile and tail quantile functions (resp.).

$$Q(1 - s) = c^{\gamma} s^{-\gamma}\left(1 + \gamma d c^{\rho} s^{-\rho} + o\left(s^{-\rho}\right)\right) \text{ as } s \to \infty, \tag{2.37}$$

and

$$U(t) = c^{\gamma} t^{\gamma}\left(1 + \gamma d c^{\rho} t^{\rho} + o(t^{\rho})\right) \text{ as } t \to \infty.$$

A simple calculation shows that in the Hall's model the functions $A(t)$ and $A^{*}(t)$ are equivalent to $d\rho\gamma c^{\rho} t^{\rho}$ and $d\rho\gamma c t^{\rho/\gamma}$ as $t \to \infty$, considering that the function $A^{**}(t)$ is equivalent to $d\rho\gamma c^{\rho} s^{-\rho}$ as $s \to \infty$.

**Theorem 2.3** (Asymptotic properties of $\hat{\gamma}^{H}_{k(n)}$)**.**
*Suppose that $F \in \mathcal{D}\left(H_{1/\gamma}\right)$ with $\gamma > 0$. Then as $n \to \infty$, $k \to \infty$ and $k/n \to 0$*

$(a)$ *Weak Consistency :*

$$\hat{\gamma}^{H}_{k(n)} \xrightarrow{p} \gamma \text{ as } n \to \infty.$$

$(b)$ *Strong consistency : If $k/\log\log n \to \infty$ as $n \to \infty$, then*

$$\hat{\gamma}^{H}_{k(n)} \xrightarrow{a.s.} \gamma \text{ as } n \to \infty.$$

$(b)$ *Asymptotic normality : Suppose that the df $F$ satisfies the second order condition (2.32). Then*

$$\sqrt{k}\left(\hat{\gamma}^{H}_{k(n)} - \gamma\right) \xrightarrow{d} \mathcal{N}\left(\frac{\lambda}{1-\rho}, \gamma^2\right) \text{ as } n \to \infty,$$

*provided $k = k(n)\sqrt{k}A(n/k) \to \lambda$ as $n \to \infty$.*

This last result allows to calculate confidence intervals for $\gamma$. For example, at a confidence level of $(1-\alpha)\,\%$, we have for $\lambda = 0$

$$\gamma \in \left[\hat{\gamma}^{H}_{k(n)} - q_{1-\alpha/2}\frac{\hat{\gamma}^{H}_{k(n)}}{\sqrt{k(n)}}; \hat{\gamma}^{H}_{k(n)} + q_{1-\alpha/2}\frac{\hat{\gamma}^{H}_{k(n)}}{\sqrt{k(n)}}\right],$$

where $q_{1-\alpha/2}$ is the quantile of order $(1-\alpha/2)$ of a standard normal distribution.

It was Mason who proved the weak consistency in [86], the strong consistency was proved in [35] by Deheuvels, Häusler and Mason, and the asymptotic normality was established in several papers such as, e.g. [30] ; [33] and [58].

**Moment Estimator**

Dekkers, Einmahl and de Haan (1989) [36] have developed as an extension of the Hill's estimator to the moment estimator which is valid whatever the sign of the index $\gamma$ not only for $\gamma > 0$ and which is defined as follows

$$\hat{\gamma}^{M}_{k(n)} := H^{(1)}_{k(n)} + 1 - \frac{1}{2}\left(1 - \frac{\left(H^{(1)}_{k(n)}\right)^2}{H^{(2)}_{k(n)}}\right)^{-1}, \tag{2.38}$$

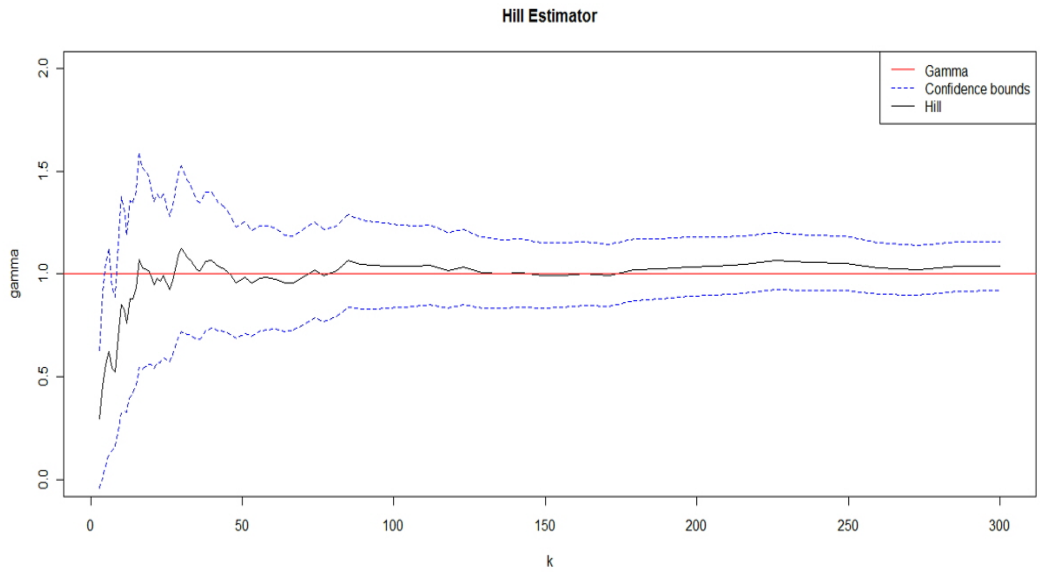



FIGURE 2.2. Hill estimator, with a confidence interval level of 95%, for the EVI of the standard Pareto distribution ($\gamma = 1$) based on 100 samples of 3000 observations.

where

$$H_{k(n)}^{(r)} := \frac{1}{k(n)} \sum_{i=1}^{k(n)} \left(\log X_{n-i+1,n} - \log X_{n-k(n),n}\right)^r, \; r = 1, 2. \tag{2.39}$$

As the moment estimator is an extension of the Hill estimator, it satisfies the asymptotic proprieties as well. The weak and strong consistency of this estimator was proved by its creators Dekkers et al (1989) [36].

**Theorem 2.4** (Asymptotic properties of $\hat{\gamma}^{M}_{k(n)}$).
*Suppose that* $F \in \mathcal{D}(H_\gamma)$, $\gamma \in \mathbb{R}$, $k \to \infty$ *and* $k/n \to 0$ *as* $n \to \infty$.

$(a)$ *Weak consistency :*

$$\hat{\gamma}^{M}_{k(n)} \xrightarrow{P} \gamma \text{ as } n \to \infty.$$

$(b)$ *Strong consistency :*

$$\hat{\gamma}^{M}_{k(n)} \xrightarrow{a.s.} \gamma \text{ as } n \to \infty.$$

$(c)$ *Asymptotic normality : (see Theorem 3.1 and corollary 3.2 of [36])*

$$\sqrt{k}\left(\hat{\gamma}^{M}_{k(n)} - \gamma\right) \xrightarrow{d} \mathcal{N}\left(0, \eta^2\right) \text{ as } n \to \infty,$$

where

$$\eta^2 := \begin{cases} 1+\gamma^2, & \gamma \geq 0, \\ \left(1+\gamma^2\right)\left(1-2\gamma\right)\left(4 - 8\dfrac{1-2\gamma}{1-3\gamma} + \dfrac{(5-11\gamma)(1-2\gamma)}{(1-3\gamma)(1-4\gamma)}\right), & \gamma < 0. \end{cases}$$

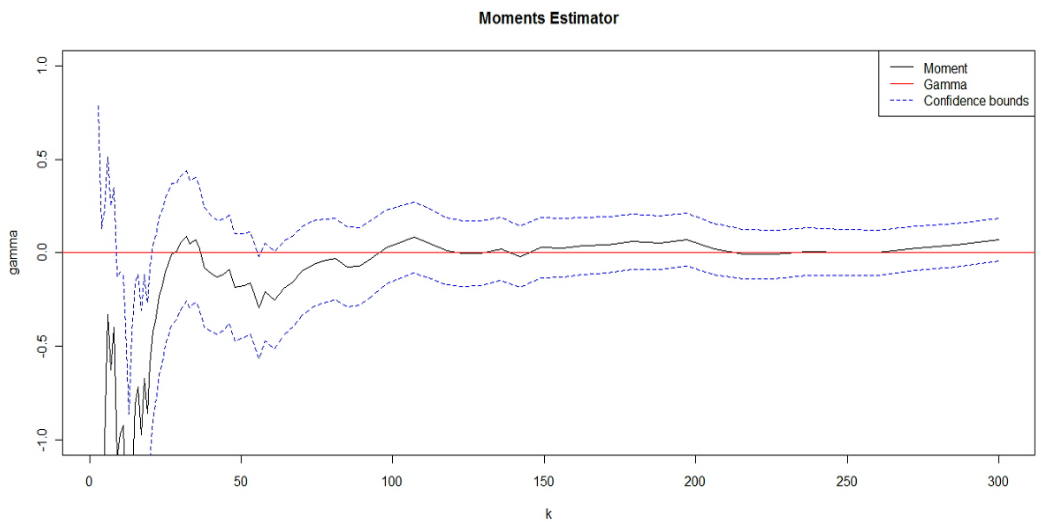



FIGURE 2.3. Estimator of the Moments, with a confidence interval level of 95%, for the EVI of the Gumbel distribution ($\gamma = 0$) based on 100 samples of 3000 observations.

**Kernel Type Estimator**

In 1985, Deuhevels, Csörgo and Mason [29] introduced a kernel-type estimator, defined as follows

$$\hat{\gamma}_{n,h}^{CDM} := \left( \sum_{i=1}^{n-1} \frac{i}{nh} K\left(\frac{i}{nh}\right) (\log X_{n-i+1,n} - \log X_{n-i,n}) \right) \left( \int_0^{1/h} K(u)\, du \right)^{-1}, \tag{2.40}$$

where $h > 0$ is called the smoothing parameter (or window) and $\{K(u) : u \geq 0\}$ is a kernel function satisfying the following conditions

$(CK1)$ $K(u) \geq 0$ for $u \in (0, \infty)$,

$(CK2)$ $K$ non-increasing and continues right on $(0, \infty)$,

$(CK3)$ $\int_0^\infty K(u)\, du = 1,$

$(CK4)$ $\int_0^\infty u^{-1/2} K(u)\, du < \infty.$

Under these conditions, the authors in [29] proved the consistency and the asymptotic normality of this estimator.

According to the choice of the kernel $K$ and the smoothing parameter $h$, different estimators can result, the best known being the Hill estimator $\hat{\gamma}_{k(n)}^{H}$, corresponding to the particular case, $K(u) = \mathbb{1}_{(0,1)}$ and $h = k/n$.

This class of estimators is valid only for $\gamma > 0$. A more general class of kernel estimators for $\gamma \in \mathbb{R}$ is given by Groeneboom, Lopuhaä and de Wolf (2003) (GLW) [52]

$$\hat{\gamma}_{n,h}^{GLW} := \hat{\gamma}_{n,h}^{(pos)} - 1 + \frac{\hat{q}_{n,h}^{(2)}}{\hat{q}_{n,h}^{(1)}}, \tag{2.41}$$

where

$$\hat{\gamma}_{n,h}^{(pos)} \quad : \quad = \sum_{i=1}^{n-1} \frac{i}{n} K_h \left( \frac{i}{n} \right) \left( \log X_{n-i+1,n} - \log X_{n-i,n} \right), \tag{2.42}$$

$$\hat{q}_{n,h}^{(1)} \quad : \quad = \sum_{i=1}^{n-1} \left( \frac{i}{n} \right)^{\alpha} K_h \left( \frac{i}{n} \right) \left( \log X_{n-i+1,n} - \log X_{n-i,n} \right), \tag{2.43}$$

$$\hat{q}_{n,h}^{(2)} \quad : \quad = \sum_{i=1}^{n-1} \frac{d}{du} \left[ u^{\alpha+1} K_h (u) \right]_{u=i/n} \left( \log X_{n-i+1,n} - \log X_{n-i,n} \right), \tag{2.44}$$

with $K_h(u) = K(u/h) h^{-1}$ and $\alpha > 0$. Here, the kernel function $K$ satisfying the following conditions

$(CK1)$ $K(u) = 0$ for $x \notin [0,1)$ and $K(u) \geq 0$ for $x \in [0,1)$,

$(CK2)$ $K$ is twice differentiable on $[0, 1[$,

$(CK3)$ $K(1) = \acute{K}(1) = 0$,

$(CK4)$ $\displaystyle\int_0^1 K(u)\, du = 1$,

$(CK5)$ For everything $\alpha > 1/2$, $\displaystyle\int_0^1 u^{\alpha-1} K(u)\, du \neq 0$.

Note that the first term of (2.41) is the kernel type estimator (almost surely) $\hat{\gamma}_{n,h}^{CDM}$. This estimator $\left(\hat{\gamma}_{n,h}^{GLW}\right)$ is based on von Mises conditions

$$\lim_{t \to x_F} \frac{d}{dt} \left( \bar{F}(t) / F'(t) \right) = \gamma, \tag{2.45}$$

where $x_F := \sup \{x : F(x) < 1\} \leq \infty$ is the set of upper limit points of $F$. The consistency of the estimator $\hat{\gamma}_{n,h}^{GLW}$ is given by the following theorem (see Theorem 3.1 in [52]).

**Theorem 2.5** (Weak consistency of $\hat{\gamma}_{n,h}^{GLW}$).
*Suppose that $F \in \mathcal{D}(H_\gamma)$, $\gamma \in \mathbb{R}$. Let us fix $\alpha > 0$ arbitrary, and either the kernel $K$ satisfies the conditions $(CK1)$-$(CK5)$. If $h = h_n, h \downarrow 0$ and $nh \to \infty$ as $n \to \infty$, then*

$$\hat{\gamma}_{n,h}^{GLW} \xrightarrow{P} \gamma \text{ as } n \to \infty.$$

*Suppose that $Q$ is differentiable, let us recall (2.45), it is convenient to write*

$$\phi(s) := -s\frac{d}{ds}\log Q(1-s), 0 < s < 1. \tag{2.46}$$

To obtain the asymptotic normality of this estimator, we require the following additional assumptions on the distribution $F$

$(CP1)$ If $\gamma \geq 0$, then $\phi(s) \to \gamma$ as $s \downarrow 0$,

$(CP2)$ If $\gamma < 0$, for a constant $c > 0$, $s^{\gamma}\phi(s) \to -c\gamma$ as $s \downarrow 0$,

$(CP3)$ If $\gamma = 0$, for all $s > 0$, $\phi(hs)/\phi(h) \to 1$ as $h \downarrow 0$.

Consider the deterministic equivalent of $\gamma_{n,h}^{(GLW)}$

$$\gamma_h^{(GLW)} := \gamma_h^{(CDM)} - 1 + \frac{q_h^{(2)}}{q_h^{(1)}}, \tag{2.47}$$

with

$$\gamma_h^{(CDM)} := \int_0^1 \log Q(1-hs)\, d(sK(s)), \tag{2.48}$$

and

$$q_h^{(i)} := h^{\alpha-1}\int_0^1 \log Q(1-hs)\, dK^{(i)}(s), i = 1, 2, \tag{2.49}$$

where

$$K^{(1)}(s) := s^{\alpha}K(s) \text{ and } K^{(2)}(s) := d\left(s^{\alpha+1}K(s)\right)/ds. \tag{2.50}$$

In the following, we use the following notation

$$x \vee y := \max(x, y) \text{ and } x \wedge y := \min(x, y), \text{ for all } (x, y) \in \mathbb{R}^2.$$

Moreover, for $\gamma \in \mathbb{R}$, we put

$$\gamma^+ := \gamma \vee 0 \text{ and } \gamma^- := \gamma \wedge 0.$$

For any function $K$ satisfying the conditions (CK2),

$$\tilde{K}(s) \; : \; = \int_s^1 t^{-1} d(tK(t)), \; s \in (0,1], \tag{2.51}$$

$$\tilde{K}^{(i)}(s) \; : \; = \int_s^1 t^{-1-(\gamma\wedge 0)} dK^{(i)}(t), \; s \in (0,1], \; i = 1, 2, \tag{2.52}$$

and

$$\bar{K}(s) := \gamma^+ \tilde{K}(s) + a_1 \tilde{K}^{(2)}(s) - a_2 \tilde{K}^{(1)}(s), \; s \in (0,1], \tag{2.53}$$

where

$$a_1 := \left( \int_0^1 t^{-1-(\gamma\wedge 0)} K^{(1)}(t)\, dt \right)^{-1} \text{ and } a_2 =: (1 + (\gamma \wedge 0))\, a_1.$$

The asymptotic normality of $\hat{\gamma}_{n,h}^{GLW}$ is given by the following Theorem.

**Theorem 2.6** (Asymptotic normality of $\hat{\gamma}_{n,h}^{GLW}$).
*Suppose that $F \in \mathcal{D}(H_\gamma)$ for $\gamma \in \mathbb{R}$, and assume that $(CP1)-(CP3)$ are satisfied. Let us fix $\alpha > 1/2$ arbitrary, and let $K$ be a kernel satisfying the conditions $(CK1) - (CK5)$. If $h = h_n$, $h \downarrow 0$ and $(nh)^{-\alpha} \log n = O\left((nh)^{-1/2}\right)$ as $n \to \infty$, we have*

$$(nh)^{1/2} \left( \hat{\gamma}_{n,h}^{GLW} - \gamma_h^{GLW} \right) \xrightarrow{d} \mathcal{N}\left(0, \sigma^2\left(\bar{K}\right)\right) \text{ as } n \to \infty,$$

*with*

$$\sigma^2\left(\bar{K}\right) = \int_0^1 \left(\bar{K}(s)\right)^2 ds. \tag{2.54}$$

Let $\mathbb{S}$ be the set of Strassen (see Strassen [111]) which consists of any absolutely continuous functions $f$ defined on $[0,1]$ such that

$$f(0) = 0 \text{ and } \int_0^1 \left(f'(s)\right)^2 ds \leq 1,$$

where $f'$ is the derivative of $f$ in the sense of Lebesgue. The following notation is used hereafter

$$l_n := \log\log(\max(n, 3)), n = 1, 2, \cdots \tag{2.55}$$

For the description of the almost sure behavior of the estimator $\hat{\gamma}_{n,h}^{GLW}$ we first give the properties of the statistics $\hat{q}_{n,h}^{(i)}, i = 1, 2$, then the estimator $\hat{\gamma}_{n,h}^{CDM}$ (for more details, see Necir [94]).

**Theorem 2.7.**
*Suppose that $F \in \mathcal{D}(H_\gamma)$ for $\gamma \in \mathbb{R}$, and assume that $(CP1)-(CP3)$ are satisfied. Let us fix $\alpha > 1/2$ arbitrary, and let $K$ be a kernel satisfying the conditions $(CK1)-(CK5)$. If $h = h_n$, $h \downarrow 0$ and $nh/l_n \to \infty$ as $n \to \infty$, with probability 1, the sequences*

$$\left\{ (nh)^{1/2} h^{1-\alpha} l_n^{-1/2} \left[ \hat{q}_{n,h}^{(i)} - q_h^{(i)} \right] \right\}_{n \geq 1}, i = 1, 2$$

*are relatively compact on $\mathbb{R}$. The sets corresponding to the endpoints are equal to*

$$\Theta\left(K^{(i)}\right) := \left\{ 2^{1/2}\gamma^+ \int_0^1 s^{-1-\gamma^-} f(s)\, dK^{(i)}(s), f \in \mathbb{S} \right\}, i = 1, 2, \tag{2.56}$$

*where $K^{(i)}(.)$, $i = 1, 2$ is that in (2.50).*

The following corollary gives a functional law of the iterated logarithm for $\hat{q}_{n,h}^{(i)}, i = 1, 2$.

**Corollary 2.1** (Functional law of the iterated logarithm for the statistics $\hat{q}_{n,h}^{(i)}$).
*Under the assumptions of Theorem 2.7, with probability 1*

$$\limsup_{n\to\infty} \pm (nh)^{1/2} h^{1-\alpha} l_n^{-1/2} \left[ \hat{q}_{n,h}^{(i)} - q_h^{(i)} \right] = 2^{1/2} \delta\left(K^{(i)}\right), i = 1, 2,$$

*where*

$$\left[\delta\left(K^{(i)}\right)\right]^2 := \left(\gamma^+\right)^2 \int_0^1 \int_0^1 \min(s, t)\, s^{-1-\gamma^-} t^{-1-\gamma^-} dK^{(i)}(s)\, dK^{(i)}(t). \tag{2.57}$$

The following corollary gives the almost sure behavior of the kernel estimator of the extreme value index $\hat{\gamma}_{n,h}^{CDM}$

**Corollary 2.2** (Strong consistency of $\hat{\gamma}_{n,h}^{CDM}$).
*Suppose that $F \in \mathcal{D}(H_\gamma)$ with $\gamma > 0$, and that $(CP1)-(CP3)$ are satisfied. Let $K$ be a kernel satisfying conditions $(CK1), (CK2)$ and $(CK5)$. If $h = h_n$ satisfies $h \downarrow 0$ and $nh/l_n \to \infty$ as $n \to \infty$, with probability $n \to \infty$*

$$\limsup_{n\to\infty} \pm (nh)^{1/2} l_n^{-1/2} \left[ \hat{\gamma}_{n,h}^{CDM} - \gamma_h^{CDM} \right] = 2^{1/2} \tilde{\delta}(K),$$

*where*

$$\left[\tilde{\delta}(K)\right]^2 := \left(\gamma^+\right)^2 \int_0^1 \int_0^1 \min(s,t)\, s^{-1} t^{-1} d\left(sK(s)\right). \tag{2.58}$$

*In addition*

$$\hat{\gamma}_{n,h}^{CDM} \overset{a.s.}{\rightarrow} \gamma^+ \text{ as } n \rightarrow \infty.$$

**Theorem 2.8** (Functional law of the iterated logarithm for the estimator $\hat{\gamma}_{n,h}^{GLW}$)**.**

*Suppose that $F \in \mathcal{D}(H_\gamma)$ for $\gamma \in \mathbb{R}$, and assume that $(CP1)-(CP3)$ are satisfied. Let us fix $\alpha > 1/2$ arbitrary, and let $K$ be a kernel satisfying the conditions $(CK1)-(CK5)$. If $h = h_n$, $h \downarrow 0$ and $nh/l_n \rightarrow \infty$ as $n \rightarrow \infty$, with probability 1, the sequence*

$$\left\{(nh)^{1/2} l_n^{-1/2} \left[\hat{\gamma}_{n,h}^{GLW} - \gamma_h^{GLW}\right]\right\}_{n\geq 1},$$

*is relatively compact on. The corresponding set of endpoints is equal to*

$$\Omega\left(\bar{K}\right) := \left\{2^{1/2} \int_0^1 \bar{K}(s)\, df(s),\, f \in \mathbb{S}\right\}, \tag{2.59}$$

*where $\bar{K}(s)$ is that in (2.53).*

**Theorem 2.9** (Strong consistency of $\hat{\gamma}_{n,h}^{GLW}$)**.**
*Under the assumption of the theorem 2.8, we have with probability* 1

$$\limsup_{n\rightarrow\infty} \pm (nh)^{1/2} l_n^{-1/2} \left[\hat{\gamma}_{n,h}^{GLW} - \gamma_h^{GLW}\right] = 2^{1/2}\sigma\left(\bar{K}\right),$$

*where $\sigma\left(\bar{K}\right)$ is that in (2.54). Furthermore*

$$\hat{\gamma}_{n,h}^{GLW} \overset{a.s.}{\rightarrow} \gamma \text{ as } n \rightarrow \infty.$$

## 2.2 POT Model Estimation Procedure

Let $(X_1, X_2, \cdots, X_n)$ be the original random sample of the rv $X$ with df $F$. Given a value of the threshold $u$, let $N_u$ be the number of exceedance of this sample. We get then, a sample of $N_u$ excesses, denoted by $Y_j = X_i - u \,|X_i > u$ for $i = 1, \cdots n$ and $j = 1, \cdots, N_u$. Suppose that the excess are iid with the GPD function. The density function $g_{\gamma,\sigma}$ of $G_{\gamma,\sigma}(x)$ is then

$$g_{\gamma,\sigma}(x) := \begin{cases} \dfrac{1}{\sigma}\left(1+\gamma\dfrac{x}{\sigma}\right)^{-1/\gamma-1} & if \quad \gamma \neq 0 \\ e^{-x/\sigma} & if \quad \gamma = 0 \end{cases}, \sigma > 0. \tag{2.60}$$

In this Section, the parametric estimation of the GPD parameters $\gamma$ and $\sigma$ will be also performed by both two methods : the Maximum Likelihood Estimation (ML) and the Probability Weighted Moments (PWM).

### 2.2.1 Maximum Likelihood Method (ML)

For $\gamma \neq 0$, the log-likelihood function, for a given random sample $(y_1, \cdots, y_{N_u})$ with GPD, can be obtained by

$$l_{\gamma,\sigma}(y_1, \cdots, y_{N_u}) := -N_u \log \sigma - \left(\frac{1}{\gamma} + 1\right) \sum_{i=1}^{N_u} \log\left(1 + \frac{\gamma}{\sigma} y_i\right), \tag{2.61}$$

where $1 + \frac{\gamma}{\sigma} y_i > 0, i = 1, \cdots, N_u$.

For $\gamma = 0$, the log-likelihood function reduce to the following expression

$$l_{0,\sigma}(y_1, \cdots, y_{N_u}) := -N_u \log \sigma - \frac{1}{\sigma} \sum_{i=1}^{N_u} y_i, \tag{2.62}$$

Taking partial derivatives of the function (2.61) with respect to $\gamma$ and $\sigma$ (resp.) the ML estimators $\left(\hat{\gamma}_{N_u}, \hat{\sigma}_{N_u}\right)$ follows then

$$\begin{cases} \dfrac{1}{N_u} \sum\limits_{i=1}^{N_u} \log\left(1 + \hat{\gamma} \dfrac{y_i}{\hat{\sigma}}\right) = \hat{\gamma}, \\ \dfrac{1}{N_u} \sum\limits_{i=1}^{N_u} \dfrac{y_i/\hat{\sigma}}{1 + \hat{\gamma} y_i/\hat{\sigma}} = \dfrac{1}{1+\hat{\gamma}}, \end{cases} \tag{2.63}$$

This method has the advantage of having good asymptotic properties, but has the disadvantage of proposing non-explicit estimators, a solution of a system of two equations with two unknowns. The latter, however, is solved by numerical algorithms.

Smith [110] shows the asymptotic normality of the ML estimators, provided $\gamma > -1/2$. Specifically we have

$$\sqrt{N_u} \begin{pmatrix} \hat{\gamma}_{N_u} - \gamma \\ \hat{\sigma}_{N_u}/\sigma_{N_u} - 1 \end{pmatrix} \xrightarrow{d} \mathcal{N}_2\left(0, (1+\gamma) \begin{pmatrix} 1+\gamma & -1 \\ -1 & 2 \end{pmatrix}\right) \text{ as } N_u \to \infty,$$

with $\mathcal{N}_2(\mu, \sum)$ stands to the bivariate normal distribution with mean vector $\mu$ and covariance matrix $\sum$. With this result, confidence intervals for the ML estimators are easily constructed.

### 2.2.2 Probability Weighted Moment Method (PWM)

Similarly to the estimation of the GEV distribution (see section 2.1.1), Hosking and Wallis [67] also suggest the use of PWM estimators for GPD. Recalling the definition of PWM estimators in 2.10, we consider for the GPD, $M_{p,r,s}$, with $p = 1, r = 0, s \in \mathbb{N}$, yielding

$$M_{1,0,s} := E\left[X\bar{G}^s_{\gamma,\sigma}(X)\right] = \frac{\sigma}{(s+1)(s+1-\gamma)} \text{ for } \gamma < 1, \tag{2.64}$$

where $X$ is a rv with df $G_{\gamma,\sigma}$. For $s = 0, 1$, we obtain

$$\gamma = 2 - \frac{M_{1,0,0}}{M_{1,0,0} - 2M_{1,0,1}} \text{ and } \sigma = \frac{2M_{1,0,0}M_{1,0,1}}{M_{1,0,0} - 2M_{1,0,1}}.$$

As for the case of a GEV distribution, we can replace $M_{1,0,s}$, for $s = 0, 1$, by its empirical estimators

$$\hat{M}_{1,0,s} := \frac{1}{N_u}\sum_{j=1}^{N_u}\left(\prod_{l=1}^{s}\frac{(N_u - j - l + 1)}{(N_u - l)}\right)Y_{j,n}, \tag{2.65}$$

yields the PWM estimators $\hat{\gamma}$ and $\hat{\sigma}$ of the GPD parameters.

### 2.2.3 Estimating Distribution Tails

Once the GPD parameters are estimated by one of the above methods. Such a formulation is given by the following equality

$$\bar{F}(x) = \bar{F}_u(x-u)\bar{F}(u), \ u < x < x_F. \tag{2.66}$$

In order to obtain an estimate for the tail $\bar{F}(x)$, the estimators of the conditional tail $\bar{F}_u(x-u)$ and $\bar{F}(u)$ are needed. After, we compute the estimates of the GPD parameters and by virtue of (1.48), the conditional tail $\bar{F}_u$ of $F$ can be estimated by

$$\widehat{\bar{F}}_u(x-u) := \bar{G}_{\hat{\gamma}_u,\hat{\sigma}_u}(x-u) = \left(1 + \hat{\gamma}_u\frac{x-u}{\hat{\sigma}_u}\right)^{-1/\hat{\gamma}_u}, \ u < x < x_F, \tag{2.67}$$

as well as $\bar{F}(u)$ is estimated by the empirical probability of exceedance

$$\widehat{\bar{F}}(u) := \bar{F}_n(u) = \frac{1}{n}\sum_{i=1}^{n}\mathbb{1}_{\{X>u\}} = \frac{N_u}{n}, \ u < x_F. \tag{2.68}$$

Putting all this together, the distribution tail estimator is therefore

$$\widehat{\bar{F}}(x) := \frac{N_u}{n}\left(1 + \hat{\gamma}_u\frac{x-u}{\hat{\sigma}_u}\right)^{-1/\hat{\gamma}_u}, \ u < x < x_F. \tag{2.69}$$

## 2.3 Optimal Sample Fraction Selection

The results concerning the estimators of the EVI stated above are asymptotic, they are obtained when $k = k(n) \to \infty$ and $k/n \to 0$, as $n \longrightarrow \infty$ (i.e. $k$ must be large enough, but not too large : it must increase moderately as the sample size increases). An important issue in semi-parametric approaches is the consideration of $k$, which represents on the one hand the quantity of data which one extracts from a sample of size $n$ for the estimation of the EVI and on the other hand the value of the threshold $u$ from which one can use the estimator (2.94). In practice, the choice of $k(n)$ is crucial for the semi-parametric estimators to have desirable properties.

For a sample of given size $n$, if $k$ is sufficiently high, the number of order statistics used increases, allowing a decrease of the estimators' variance but resulting in a larger bias. On the other hand, if $k$ is sufficiently low, we stay close to the sample maximum and few order statistics will be used, resulting in estimators with large variances. It is therefore necessary to make a compromise between bias and variance to obtain the optimal value $k_{opt}$ of $k$ (and in an equivalent way the optimal value $u_{opt}$ of $u$). Several methods have been proposed, we present in this section most of them.

### 2.3.1 Graphical Method

One of the most used methods in practice is the Hill plot. This is a heuristic approach. Given a sample of size $n$, we plot the Hill estimator for different choices of $k$, i.e. the graph

$$\left\{ \left(k, \hat{\gamma}^H_{k(n)}\right) : k = 2, \cdots, n \right\}, \tag{2.70}$$

and retains the value $k_{opt}$ that correspond to a reasonable horizontal plot, is considered for an elective value of the estimator $\hat{\gamma}^H_{k(n)}$.

### 2.3.2 Minimization of the Asymptotic Mean Square Error

An important criterion, very popular among statisticians and used in most of the articles, is choosing $k$ in order to minimize the Asymptotic Mean Squared Error (AMSE). However, this criterion is mainly dependent on the second order assumptions about the underlying df $F$. The value $k_{opt}$ can be determined when the analytical form of $F$ is known (or estimated by the bootstrapping methods).

For an arbitrary estimator of $\hat{\gamma}_{k(n)}$, we define the AMSE as follows:

$$AMSE\left(\hat{\gamma}_{k(n)}\right) := E_{\infty}\left[\left(\hat{\gamma}_n-\gamma\right)^2\right] = AVar\left(\hat{\gamma}_n\right) + ABias^2\left(\hat{\gamma}_n\right), \tag{2.71}$$

where $E_{\infty}$ denotes the asymptotic mean value, $AVar$ and $ABias$ stands for Asymptotic Variance and Asymptotic Bias, resp. Then the idea is to choose an optimal sequence $k_{opt}(n)$ which minimizes the AMSE. It is therefore to choose

$$k_{opt} := \arg\min_{k} AMSE\left(\hat{\gamma}_{k(n)}\right),$$

### 2.3.3 Adaptive Procedures

In the literature, several adaptive methods were developed which we review briefly for the choice of the number of extreme order statistics of $k$, for special classes of distributions.

#### Hall and Welsh Approach

Hall and Welsh [62] have shown that if the cdf $F$ satisfies the Hall condition 2.36, then the asymptotic mean square error (AMSE) of the Hill estimator $\hat{\gamma}^H_{k(n)}$ is minimal for

$$k_{opt} \sim \left(\frac{c^{2\rho}\left(1+\rho\right)^2}{2d^2\rho^3}\right)^{1/(2\rho+1)} n^{2\rho/(2\rho+1)} \text{ as } n \to \infty, \tag{2.72}$$

However, this result can not be used directly to determine the optimal number of order statistics because the parameters $\rho, c$ and $d$ are unknown. Hall and Welsh [62] constructed a consistent estimator for $k_{opt}$

$$\hat{k}_{opt} := \left[\hat{\delta} n^{2\hat{\rho}/(2\hat{\rho}+1)}\right], \tag{2.73}$$

where

$$\hat{\rho} := \left|\log\left|\frac{\left(\hat{\gamma}^H_{k(n)}(t_1)\right)^{-1} - \left(\hat{\gamma}^H_{k(n)}(s)\right)^{-1}}{\left(\hat{\gamma}^H_{k(n)}(t_2)\right)^{-1} - \left(\hat{\gamma}^H_{k(n)}(s)\right)^{-1}}\right| / \log\left(\frac{t_1}{t_2}\right)\right|, \tag{2.74}$$

and

$$\hat{\delta} := \left|(2\hat{\rho})^{-1/2}\left(\frac{n}{t_1}\right)^{\hat{\rho}} \frac{\left(\hat{\gamma}^H_{k(n)}(t_1)\right)^{-1} - \left(\hat{\gamma}^H_{k(n)}(t_3)\right)^{-1}}{\hat{\gamma}^H_{k(n)}(s)}\right|^{2/(2\hat{\rho}+1)}, \tag{2.75}$$

in the sense that $\frac{\hat{k}_{opt}}{k_{opt}} \xrightarrow{P} 1$ as $n \to \infty$, with $t_i = [n^{\tau_i}], i = 1, 2$ and $s = [n^{\sigma}]$ for some $0 < 2\rho(1-\tau_1) < \sigma < 2\rho/(2\rho+1) < \tau_1 < \tau_2 < 1$; $[x]$ denotes here the largest integer less than or equal to $x$.

**Bootstrap Approach**

A new resampling procedure to select the number of extreme order statistics, through the mean squared error of the Hill estimator, is proposed. For this purpose, the usual bootstrap does not work properly, especially because it seriously underestimates bias. To circumvent this problem, Hall [60] proposes to use resamples of smaller size than the original one and linking the bootstrap estimates for the optimal subsample fraction to $k_{opt}$ for the full sample. However, in order to establish this link, Hall's method requires that $\rho = 1$, which puts a serious restriction on the tail behavior of the data.

Recently, the idea of subsample bootstrapping is taken up in a broader method by Danielsson et al [31]. They used a combination of the subsample bootstrapping estimates for the difference of two estimators based on bootstrap samples of different order to obtain a convergent estimator of the optimal number of order statistics that requires no restrictions on $\rho$. Draisma, de Haan and Peng [41] have developed a method based on a double bootstrap. They are concerned with the more general case $\gamma \in \mathbb{R}$, and their results relate to Pickands and moments estimators.

**Sequential Approach**

Drees and Kaufmann [42] present a sequential approach to select the optimal sample fraction. From a law of the iterated logarithm, they construct "stopping times" for the sequence of Hill estimators that are asymptotically equivalent to a deterministic sequence

$$\tilde{k}(r) := \min\left\{k \in \{2, \cdots, n\} : \max_{2 \le i \le k} i^{1/2} \left|\hat{\gamma}^H_{k(n)}(i) - \hat{\gamma}^H_{k(n)}(k)\right| > r\right\}, \tag{2.76}$$

where $r = r_n$ constitute a sequence which is of higher order than $(\log \log n)^{1/2}$ and of lower order than $n^{1/2}$. In comparison with other procedures, the influence of a wrong specification of the parameter $\rho$ in this method does not seem to be a major problem. The sequential procedure give the best results even when setting $\rho = 1$. The Drees and Kaufmann method can be described by the following algorithm :

**Step 1 :** For $r = 2.5 \times n^{0.25} \times \hat{\gamma}_{k(n)}$ with an initial estimate $\hat{\gamma}_{k(n)} := \hat{\gamma}^{H}_{k(n)}\left(2\sqrt{n}\right)$.

**Step 2 :** Compute the "stopping time"

$$\tilde{k}(r) := \min\left\{ k \in \{1, \cdots, n-1\} \,\middle|\, \max_{1 \leq i \leq k} \sqrt{i}\left|\hat{\gamma}^{H}_{k(n)}(i) - \hat{\gamma}^{H}_{k(n)}(k)\right| > r \right\}.$$

**Step 3 :** Similarly, compute $\tilde{k}\left(r^{\xi}\right)$ for $\xi = 0.7$.

**Step 4 :** Calculate

$$\hat{k}_{opt} := \left[ \frac{1}{3}\left(2\left(\hat{\gamma}^{H}_{k(n)}\right)^{2}\right)\left(\frac{\tilde{k}\left(r^{\xi}\right)}{\left(\tilde{k}(r)\right)^{\xi}}\right)^{1/(1-\xi)} \right],$$

where $[x]$ denotes the largest integer less than or equal to $x$. For more details on this approach, we refer to [42], [95] et [8]

### Cheng and Peng Approach

The asymptotic normality of Hill's estimator is used to construct confidence intervals for the EVI $\gamma$ of a cdf $F$ belonging to Hall's class. Indeed, we have the following proposition (see, e.g. Hall, 1982 [61])

**Proposition 2.2.**
*Suppose that (2.36) is satisfied and $k \to \infty$, $k/n \to 0$. Then*

$$\sqrt{k}\left(\hat{\gamma}^{H}_{k(n)} - \gamma\right) \xrightarrow{d} \mathcal{N}\left(0, \gamma^{2}\right) \quad \text{as } n \to \infty, \tag{2.77}$$

*iff* $k = o\left(n^{-2\rho/(1-2\rho)}\right)$.

Thus, for $0 < \alpha < 1$ the one-sided and two-sided intervals of confidence level $(1 - \alpha)$ for the EVI $\gamma$ are (resp.)

$$I_1(\alpha) := \left(0, \hat{\gamma}^{H}_{k(n)} + z_{\alpha}\frac{\hat{\gamma}^{H}_{k(n)}}{\sqrt{k}}\right), \tag{2.78}$$

and

$$I_2(\alpha) := \left( \hat{\gamma}^H_{k(n)} - z_{\alpha/2} \frac{\hat{\gamma}^H_{k(n)}}{\sqrt{k}}, \hat{\gamma}^H_{k(n)} + z_{\alpha/2} \frac{\hat{\gamma}^H_{k(n)}}{\sqrt{k}} \right), \tag{2.79}$$

where $z_\omega$ $(0 < \omega < 1)$ is the $(1-\omega)$-quantile of the standard normal distribution, defined by $P(\mathcal{N}(0,1) \leq z_\omega) = 1 - \omega$. It is shown in Cheng and Peng (2001) [25] that, as $k \to \infty$ and $k/n \to 0$, the corresponding coverage probabilities are

$$\begin{aligned} & P(\gamma \in I_1(\alpha)) = \\ & 1 - \alpha - \phi(z_\alpha) \left\{ \frac{1 + 2z_\alpha^2}{3\sqrt{k}} - \frac{\rho d c^\rho}{(1-\rho)} \sqrt{k} \left( \frac{n}{k} \right)^\rho \right\} + o\left( \frac{1}{\sqrt{k}} + \sqrt{k} \left( \frac{n}{k} \right)^\rho \right), \end{aligned} \tag{2.80}$$

and

$$P(\gamma \in I_2(\alpha)) = 1 - \alpha + o\left( \frac{1}{\sqrt{k}} + \sqrt{k} \left( \frac{n}{k} \right)^\rho \right), \tag{2.81}$$

with $\phi(.)$ is the density of the standard normal distribution.

By minimizing the absolute coverage error for $I_1(\alpha)$, Cheng and Peng (2001) [25] propose an optimal sample fraction

$$k_{opt} := \begin{cases} \left( \dfrac{(1+2z_\alpha^2)(1-\rho)^{1/(1-\rho)}}{-3dc^\rho \rho (1-2\rho)} \right)^{1/(1-\rho)} n^{-\rho/(1-\rho)} & if \quad d > 0, \\ \left( \dfrac{(1+2z_\alpha^2)(1-\rho)}{3dc^\rho \rho} \right)^{1/(1-\rho)} n^{-\rho/(1-\rho)} & if \quad d < 0, \end{cases} \tag{2.82}$$

Notice that it is readily verified that $k = o\left(n^{-2\rho/(1-2\rho)}\right)$ in the proposition 2.2. Since $k_{opt}$ depends on quantities characterizing the unknown cdf $F$, Cheng and Peng [25] introduce a plug-in estimate

$$\hat{k}_{opt} := \begin{cases} \left( \dfrac{1+2z_\alpha^2}{3\hat{\delta}(1-2\hat{\rho})} \right)^{1/(1-\hat{\rho})} n^{-\hat{\rho}/(1+\hat{\rho})} & if \quad \hat{\rho} > 0, \\ \left( \dfrac{1+2z_\alpha^2}{-3\hat{\delta}} \right)^{1/(1-\hat{\rho})} n^{-\hat{\rho}/(1-\hat{\rho})} & if \quad \hat{\rho} < 0. \end{cases} \tag{2.83}$$

where

$$\hat{\rho} := -(\log 2)^{-1} \log \left( \left| \frac{H^{(2)}_{k(n)}\left(n/\left(2\sqrt{\log n}\right)\right) - 2\left[H^{(1)}_{k(n)}\left(n/\left(2\sqrt{\log n}\right)\right)\right]^2}{H^{(2)}_{k(n)}\left(n/\sqrt{\log n}\right) - 2\left[H^{(1)}_{k(n)}\left(n/\sqrt{\log n}\right)\right]^2} \right| \right), \tag{2.84}$$

and

$$\hat{\delta} := (1 - \hat{\rho}) (\log n)^{-\hat{\rho}/2} \frac{H_{k(n)}^{(2)} \left(n/\sqrt{\log n}\right) - 2 \left\{ H_{k(n)}^{(1)} \left(n/\sqrt{\log n}\right) \right\}^2}{-2\hat{\rho} \left[ H_{k(n)}^{(1)} \left(n/\sqrt{\log n}\right) \right]^2}, \tag{2.85}$$

with $H_{k(n)}^{(r)}$ is given by equation (2.39).

**Reiss and Thomas Approach**

Reiss and Thomas [103, page 137] have proposed a heuristic method very simple to implement. It is enough to choose automatically for $k_{opt}$ as the value $k$ that minimizes

$$\frac{1}{k} \sum_{i \leq k} i^{\beta} \left| \hat{\gamma}_n (i) - med \left( \hat{\gamma}_n (1), \cdots \hat{\gamma}_n (k) \right) \right|, \ 0 \leq \beta \leq 1/2, \tag{2.86}$$

where $\hat{\gamma}_n (i)$ is an estimator of $\gamma$ based on the $i$ largest values of a sample of size $n$ and $med \left( \hat{\gamma}_n (1), \cdots \hat{\gamma}_n (k) \right)$ denotes the median of $\hat{\gamma}_n (1), \cdots \hat{\gamma}_n (k)$.

Reiss and Thomas have also suggested minimizing the following modification

$$\frac{1}{k-1} \sum_{i<k} i^{\beta} \left( \hat{\gamma}_n (i) - \hat{\gamma}_n (k) \right)^2, \ 0 \leq \beta \leq 1/2. \tag{2.87}$$

For a discussion on the choice of $\beta$, one refers to paper of Neves and Fraga Alves [95].

### 2.3.4 Threshold Selection

The choice of the threshold $u$ is still an unsolved problem and in the literature of the POT method, not so much attention has been given to this issue. It remains equivalent to that of the number $k$ The choice of such a threshold is subject to a trade-off between high values of $u$ (too few exceedance), where the bias of the estimators is smaller, and low values of $u$ (too many exceedance), where the variance is smaller.

For this purpose, Davison and Smith [34] suggest the use of the plot of the mean-excess function (mef), called mef-plot, defined as follow.

**Definition 2.2** (mef-plot).
*The mef-plot is given by*

$$\{(u, e_n(u)), X_{1,n} < u < X_{n,n}\}, \tag{2.88}$$

*where* $e_n(u)$ *is the empirical estimator of the mef defined in (1.49)*

$$e_n(u) := \frac{1}{\bar{F}_n(u)} \int_u^{\infty} \bar{F}_n(x) = \frac{1}{N_u} \sum_{i=1}^{n} (X_i - u) \mathbb{I}_{\{X_i > u\}}, \tag{2.89}$$

*with* $N_u$ *the number of observations exceeding* $u$*.*

Therefore, we have to check the linearity of the plot above and choose $u$ such that $e_n(x)$ is approximately linear for $x \geq u$. In other words, the threshold $u$ is chosen at the point to the right of which a rough linear pattern appears in the plot. Thus, the slope of the plot leads to a quick estimate of $\gamma$: in particular, an increasing plot indicates $\gamma > 0$, a decreasing plot indicates $\gamma < 0$, and one of roughly constant slope indicates that $\gamma$ is near 0.

Another procedure consists in choosing the $(k+1)$th largest observation $X_{n-k,n}$ as a threshold, the problem becomes a matter of which value of $k$ to take as an optimal choice.

## 2.4 Estimating High Quantiles

High quantile estimation plays an important role in the context of risk management where it is crucial to evaluate adequately the risk of a great loss what occurs very rarely.

For $0 < p < 1$, the $(1-p)$-quantile denoted by $x_p$, of the continuous strictly increasing df $F$, is defined as the solution of equation

$$F(x_p) = 1 - p.$$

If $p$ is fixed, then an estimator of $x_p$ is the empirical quantile $X_{(n-[pn],n)}$. The problem is to estimate the $(1-p)$-quantile, when $p$ is close to 0. As we use asymptotic theory, $p$ must depend on the sample size $n$ (i.e. $p := p_n$). So we are looking for an estimator of $x_{p_n}$ when $np_n$ has a limit $c$ as $n \to \infty$. We say that the $(1-p_n)$-quantile is within the sample if $c > 1$, and that the $(1-p_n)$-quantile is outside the data if $c < 1$. We refer to [89].

### 2.4.1 GEV Distribution Based Estimators

Motivated by Theorem 1.13, the GEV provides a model for the distribution of extremes of a series of independent observations $X_1, X_2, \cdots$. Data are blocked into sequences of observations of equal length $n$, for some large value of $n$, generating a series of block maxima, $M_{n,1}, ..., M_{n,m}$, say, to which the GEV distribution can be fitted. Often the blocks are chosen to correspond to a time period of length one year, in which case $n$ is the number of observations in a year and the block maxima are annual maxima. The extreme quantiles estimators $\left(\hat{x}_p = H_{\hat{\theta}}^{\leftarrow}(1-p)\right)$ of the annual maximal distribution (i.e. the $(1-p)$-quantiles) can be then obtained by inverting the distribution function $H_\theta$ given by (2.1) and replacing $\theta = (\gamma, \mu, \sigma)$ by $\hat{\theta} = (\hat{\gamma}, \hat{\mu}, \hat{\sigma})$ either obtained by the ML or PWM estimators. That

$$\hat{x}_p := \begin{cases} \hat{\mu} - \dfrac{\hat{\sigma}}{\hat{\gamma}}\left(1 - (-\log(1-p))^{-\hat{\gamma}}\right) & if \quad \gamma \neq 0, \\ \hat{\mu} - \hat{\sigma}\log(-\log(1-p)) & if \quad \gamma = 0. \end{cases} \tag{2.90}$$

In the case where $\gamma < 0$, the endpoint is finite and it can be estimated by

$$\hat{x}_F := \hat{\mu} - \frac{\hat{\sigma}}{\hat{\gamma}}. \tag{2.91}$$

In common terminology, $\hat{x}_p$ is the return level associated with the return period $T = 1/p$, since to a reasonable degree of accuracy, the level $\hat{x}_p$ is expected to be exceeded on average once every $1/p$ years. More precisely, $\hat{x}_p$ is exceeded by the annual maximum in any particular year with probability $p$, for more details see e.g. coles (2001) [28].

#### Case where F belongs to the Domain of Attraction of $H_\theta$

Using relation (1.36) with large threshold $u = a_n x + b_n$, we obtain a tail estimator of the form

$$\widehat{\overline{F}}(u) = \frac{1}{n}\left(1 + \hat{\gamma}\frac{u - \hat{b}_n}{\hat{a}_n}\right)^{-1/\hat{\gamma}}, \tag{2.92}$$

where$\hat{\gamma}$, $\hat{a}_n$ and $\hat{b}_n$ are appropriate estimates (based on the $k$ upper order statistics) of the tail index $\gamma$, and the normalizing constants $a_n$ and $b_n$ (resp.). In the case where the extreme quantiles are within the data (i.e. $p \geq 1/n$). It can be estimated by

$$\hat{x}_p := \hat{a}_n \frac{(np)^{-\hat{\gamma}} - 1}{\hat{\gamma}} + \hat{b}_n. \tag{2.93}$$

The normalizing constants $\hat{a}_n$ and $\hat{b}_n$ have a very large variance, because they are based on high quantiles of $X$. To solve this problem (Dekkers and de Haan [37]; Dekkers et al. [36]) propose to use the larger values $k$ of the sample to estimate the tail of the distribution. For $x$ sufficiently large

$$\widehat{\overline{F}}(u) = \frac{k}{n}\left(1 + \hat{\gamma}\frac{u - \hat{b}_{n/k}}{\hat{a}_{n/k}}\right)^{-1/\hat{\gamma}}, \tag{2.94}$$

We thus deduce the most typical case where the extreme quantiles are outside the data (i.e. $p < 1/n$)

$$\hat{x}_p := \hat{a}_{n/k}\frac{(np/k)^{-\hat{\gamma}} - 1}{\hat{\gamma}} + \hat{b}_{n/k}. \tag{2.95}$$

When $\gamma < 0$, the end point is finite and it can be estimated by

$$\hat{x}_F := \hat{b}_{n/k} - \frac{\hat{a}_{n/k}}{\hat{\gamma}}. \tag{2.96}$$

The extreme quantile estimators associated to the semi parametric estimators that we present are written in this form. It is therefore necessary to give estimates for the normalizing constants $a_{n/k}$ and $b_{n/k}$.

The estimator of the $(1-p)$-quantile linked to Pickands' estimator is of the following form

$$\hat{x}_p^P := X_{n-k+1,n} + \frac{(np/k)^{-\hat{\gamma}^P_{k(n)}} - 1}{1 - 2^{-\hat{\gamma}^P_{k(n)}}}\left(X_{n-k+1,n} - X_{n-2k+1,n}\right). \tag{2.97}$$

where $\hat{a}_{n/k} = \frac{\hat{\gamma}^P_{k(n)}}{1-2^{-\hat{\gamma}^P_{k(n)}}}\left(X_{n-k+1,n} - X_{n-2k+1,n}\right)$ and $\hat{b}_{n/k} = X_{n-k+1,n}$. The asymptotic properties of this estimator are discussed in Dekkers & de Haan (1989) [37].

When $\gamma < 0$, the endpoint is finite. It can be estimated by

$$\hat{x}_F^P := X_{n-k+1,n} + \frac{\left(X_{n-k+1,n} - X_{n-2k+1,n}\right)}{2^{-\hat{\gamma}^P_{k(n)}} - 1}. \tag{2.98}$$

For the Fréchet class ($\gamma > 0$), the classical Weissman type estimator of the $(1-p)$-quantile takes on the following form

$$\hat{x}_p^W := X_{n-k,n}\left(\frac{k}{np}\right)^{\hat{\gamma}^H_{k(n)}}, \tag{2.99}$$

where $\hat{\gamma}^H_{k(n)}$ is the Hill's estimator and $\hat{b}_{n/k} = \hat{a}_{n/k}/\hat{\gamma}^H_{k(n)} = X_{n-k,n}$.

The quantile of order $(1-p)$ on the basis of the moment estimator is

$$\hat{x}_p^M := X_{n-k,n} + \frac{\left(\frac{np}{k}\right)^{-\hat{\gamma}_{k(n)}^M} - 1}{\hat{\gamma}_{k(n)}^M} \frac{X_{n-k,n} M_{k(n)}^{(1)}}{\rho\left(\hat{\gamma}_{k(n)}^M\right)}, \tag{2.100}$$

where $M_{k(n)}^{(1)}$ is equal to $H_{k(n)}^{(1)}$ defined by (2.39) and

$$\rho\left(\hat{\gamma}_{k(n)}^M\right) := \begin{cases} 1, & \gamma \geq 0, \\ \dfrac{1}{1-\gamma} & \gamma < 0. \end{cases} \tag{2.101}$$

In that case $\hat{a}_{n/k} = \dfrac{X_{n-k,n} M_{k(n)}^{(1)}}{\rho\left(\hat{\gamma}_{k(n)}^M\right)}$ and $\hat{b}_{n/k} = X_{n-k,n}$.

As $\gamma < 0$, the endpoint is finite and it can be estimated by

$$\hat{x}_F^M := X_{n-k,n} + \left(1 - 1/\hat{\gamma}_{k(n)}^M\right) X_{n-k,n} M_{k(n)}^{(1)}. \tag{2.102}$$

### 2.4.2 Estimators Based on the POT Models

For fixed threshold $u$, an estimator of quantiles $x_p > u$ is obtained by inverting the expression of the tail estimate formula (2.69)

$$\hat{x}_p := u + \frac{\hat{\sigma}_u}{\hat{\gamma}_u}\left(\left(\frac{N_u}{np}\right)^{\hat{\gamma}_u} - 1\right), \; p < \frac{N_u}{n}, \tag{2.103}$$

with $\hat{\sigma}_u$ and $\hat{\gamma}_u$, the estimators of the parameters of the GPD and $N_u$ the number of excesses.

This expression can be found, e.g. in Davison and Smith (1990) [34] and Embrechts et al. (1997) [47]

The endpoint of the distribution, as $\gamma < 0$, is also estimated by

$$\hat{x}_F := u - \frac{\hat{\sigma}_u}{\hat{\gamma}_u}. \tag{2.104}$$

The threshold $u$ is often chosen equal to one of the order statistics $X_{1,n} \leq X_{2,n} \leq \cdots \leq X_{n,n}$. If one chooses as a threshold $u = X_{n-k,n}$ the $(k+1)$th largest observation, then $N_u = k$ and the high order quantile estimator is rewritten as follows

$$\hat{x}_p^{(POT)} := X_{n-k,n} + \frac{\hat{\sigma}^{(POT)}}{\hat{\gamma}^{(POT)}}\left(\left(\frac{k}{np}\right)^{\hat{\gamma}^{(POT)}} - 1\right), \; p < \frac{k}{n}, \tag{2.105}$$

where $\hat{\gamma}^{(POT)}$, $\hat{\sigma}^{(POT)}$ are the resulting estimators of $\gamma$ and $\sigma$ (resp.).

The endpoint is therefore estimated by

$$\hat{x}_F^{(POT)} := X_{n-k,n} + \frac{\hat{\sigma}^{(POT)}}{\hat{\gamma}^{(POT)}}. \tag{2.106}$$

## 2.5 Risk Measurement

Assessing the probability of rare and extreme events is an important issue in the risk management of portfolios. Extreme value theory provides the solid fundamentals needed for the statistical modelling of such events and the computation of extreme risk measures. These section is devoted to the theoretical description of risk and the risk measures. We first present a theoretical framework within which we define the concept of coherence. We than refer to some risk measures such as the Value-at-Risk, Expected Shortfall and the return level,etc.

Since 1997 the paper of Artzner et al [3] risk measurement, and hence risk measures, have gained enormously in interest under economist, bank regulators and mathematicians, giving rise to a new theory. A good reference for the Risk theory is the book of Denuit et al [38] and Kaas et al [75], see also [20].

### 2.5.1 Definitions

**Definition 2.3** (Risk).
*Risk is the future net worth of a position*

At the beginning risk measurement was mainly focussed on the mathematical properties which reflect the underlying economical meaning, however in the last years the statistical properties have become of increasing interest. Nowadays it is obvious to all working with risk, be it in practice or theory, that the procedure of risk measurement in fact involves two steps.

**1)** Estimating the loss distribution of the position.

**2)** Constructing a risk measure that summarizes the risk of the position.

The position's loss distribution in practice is generally unknown, and therefore must be estimated from data. The estimation is essentially done by backtesting.

Recall that backtesting is the procedure of periodically comparing the forecasted risk measure with realized values in the financial market. Each one of the steps above should be regarded as equally important. Because risk measurement is of great practical importance, risk measures should be formalized with the regulations of the practical world in mind. For this reason risk measures are mostly considered to be single valued. Taking a risk to be a single value can be problematic however, for instance a single number does not give any information about which risk within the position is problematic. But this is only the case when a risk is found to be unacceptable, than the portfolio should be rebalanced. If on the other hand the risk is found to be acceptable, these sort of problems do not play any part. Thus in this setting taking a single valued risk measure is justified.

Since risks are modelled as non-negative random variables rv's, measuring risk is equivalent to establishing a correspondence between the space of rv's and non-negative real numbers $\mathbb{R}^+$. The real number denoting a general risk measure associated with the risk $X$ will henceforth be denoted as $\rho(X)$. Thus, a risk measure is nothing but a functional that assigns a non-negative real number to a risk.

It is essential to understand which aspect of the riskiness associated with the uncertain outcome the risk measure attempts to quantify. No risk measure can grasp the whole picture of the danger inherent in some real-life situation, but each of them will focus on a particular aspect of the risk. There is a parallel with mathematical statistics, where characteristics of distributions may have quite different meanings and uses – for example, the mean to measure central tendency, the variance to measure spread, the skewness to reflect asymmetry and the peakedness to measure the thickness of the tails.

In this section, we will concentrate on risk measures that measure upper tails of distribution functions. We are now ready to state the definition of a risk measure.

**Definition 2.4** (Risk measure)**.**
*A risk measure is defined as a functional $\rho$ mapping a risk $X$ from the set of random variables namely losses or payments, to the set of non-negative real numbers, possibly infinite, representing the extra cash which has to be added to make it acceptable.*

The idea is that $\rho$ quantifies the riskiness of $X$ : large values of $\rho(X)$ tell us that the risk is dangerous. Specifically, if $X$ is a possible loss of some portfolio over a time horizon, we interpret $\rho(X)$ as the amount of capital that should be added as a buffer to this portfolio so that it becomes acceptable to an internal or external risk controller. In such a case, $\rho(X)$ is the risk capital of the portfolio. Such risk measures are used for determining provisions and capital requirements in order to avoid insolvency.

### 2.5.2 Premium Calculation Principles

Risk measures are in many respects akin to actuarial premium calculation principles. For an insurance company exposed to a liability $X$, a premium calculation principle $\rho$ gives the minimum amount $\rho(X)$ that the insurer must raise from the insured in order that it is in the insurer's interest to proceed with the contract. Premium principles are thus prominent examples of possible risk measures. Their characteristic is that the number resulting from their application to some insurance risk $X$ is a candidate for the premium associated with the contract providing coverage against $X$.

Premium principles are the most common risk measures in actuarial science. Although there is a consensus (at least if everyone agrees on the risk distribution) about the net premium (which is the expected claim amount), there are many ways to add a loading to it to get the gross premium. The safety loading added to the expected claim cost by the company reflects the danger associated to the risk borne by the insurer. Premium calculation principles are thus closely related to risk measures. Indeed, those principles have to express the insurer's feelings about the risk he bears. The premium for a less attractive risk should exceed the premium for a more attractive risk. Therefore, a premium calculation principle is a particular case of a risk measure.

#### Proprieties of Premium Calculation Principle

The risk measures have to satisfy certain axioms, such as those discussed in this section. The choice of a premium principle depends heavily on the importance attached to such properties. There is no premium principle that is uniformly best.

First, we present some notation that we use throughout this section. Let $\mathcal{X}$ denote the set of non-negative rv's on the probability space $(\Omega, F, P)$ ; this our collection of insurance-loss random variables, also called insurance risks. Let $X, Y, Z$,etc. denote typical members of $\mathcal{X}$. Finally, let $\rho$ denote the premium principle or function, from $\mathcal{X}$ to the set of non-negative real numbers.

**1. Independence**

$\rho[X]$ depends only on the df of $X$, namely $S_X$, in which

$$S_X(t) = P\{\omega \in \Omega : X(\omega) > t\}. \tag{2.107}$$

That is, the premium depends only on the tail probabilities of $X$: This property states that the premium depends only on the monetary loss of the insurable event and the probability that a given monetary loss occurs, not the cause of the monetary loss.

**2. Risk loading**

Loading for risk is desirable because one generally requires a premium rule to charge at least the expected payout of the risk $X$, namely $\mathbb{E}(X)$, in exchange for insuring the risk. Otherwise, the insurer will lose money on average.

$$\rho(X) \geq \mathbb{E}(X) \text{ for all } X \in \mathcal{X}. \tag{2.108}$$

**3. No unjustified risk loading**

If a risk $X$ is identically equal to a constant $c \geq 0$ (almost everywhere), then

$$\rho(X) = c. \tag{2.109}$$

In contrast to Property 2, if we know for certain (with probability 1) that the insurance payout is $c$, then we have no reason to charge a risk loading because there is no uncertainty as to the payout.

**4. Maximal loss**

$$\rho(X) \leq \max(X), \text{ for all } X \in \mathcal{X}. \tag{2.110}$$

**5. Translation invariance**

$$\rho\left(X+a\right)=\rho\left(X\right)+a, \text{ for any } X\in\mathcal{X} \text{ and for any } a\geq 0. \tag{2.111}$$

If we increase a risk $X$ by a fixed amount $a$, then Property 5 states that the premium for $X+a$ should be the premium for $X$ increased by that fixed amount $a$. Otherwise, translation invariance suggests that adding safe capital to a financial position, decreases the riskiness of the position by the same amount. This property again suggests the idea of risk measure as capital requirement, as $\rho\left(X\right)$ represents the amount of money that, added to the financial position $X$, make it marginally acceptable.

**6. Positive homogeneity**

$$\rho\left(\lambda X\right)=\lambda\rho\left(X\right), \text{ for all } X\in\mathcal{X} \text{ and all } \lambda\geq 0. \tag{2.112}$$

Positive homogeneity implies that the risk of a payoff increases linearly with the size of the investment. Simply states that increase the position size of a portfolio will raise its risk proportionally. Thus, it reflects the possible situation where no netting or diversification occurs. In particular, a government or an exchange does not prevent many firms or investors from all taking the same position. It also implies the normalization property, that is $\rho\left(0\right)=\rho\left(0Y\right)=0\rho\left(Y\right)=0$, which is usually considered a natural condition to require.

Together with the translation invariance and monotonicity requirements, normalization allows propriety 3. Indeed, for a fixed capital $X=c$, then $\rho\left(X\right)=\rho\left(c\right)=\rho\left(0+c\right)=\rho\left(0\right)+c=0+c=c$.

Holding some safely invested capital is of course not risky and $c$ is the maximal amount of capital that can be withdrawn maintaining the position acceptable. In most of the situations normalization is required even if positive homogeneity does not hold.

**7. Additivity**

This Property is a stronger form of Property 6. One can use a similar no-arbitrage argument to justify the additivity property

$$\rho\left(X+Y\right)=\rho\left(X\right)+\rho\left(Y\right), \text{ for all } X,Y\in\mathcal{X}. \tag{2.113}$$

**8. Subadditivity**

$$\rho(X+Y) \leq \rho(X) + \rho(Y), \text{ for all } X, Y \in \mathcal{X}. \tag{2.114}$$

The subadditivity property requires that adding two positions together should decrease the total risk. One example that is consistent with this logic is that, if an individual wishes to take the risk $X+Y$, opening two accounts separately will not help him save margin requirement of an exchange.

**9. Superadditivity**

$$\rho(X+Y) \geq \rho(X) + \rho(Y), \text{ for all } X, Y \in \mathcal{X}. \tag{2.115}$$

Superadditivity might be a reasonable property of a premium principle if there are surplus constraints that require that an insurer charge a greater risk load for insuring larger risks.

**10. Additivity for independent risks**

$$\rho(X+Y) = \rho(X) + \rho(Y), \tag{2.116}$$

for all $X, Y \in \mathcal{X}$ such that $X$ and $Y$ are independent. Some actuaries might feel that Property 7 is too strong and that the no-arbitrage argument only applies to risks that are independent. They, thereby, avoid the problem of surplus constraints for dependent risks.

Next, we consider properties of premium rules that require that they preserve common ordering of risks.

**11. Additivity for comonotonic risks**

$$\rho(X+Y) = \rho(X) + \rho(Y), \tag{2.117}$$

for all $X, Y \in \mathcal{X}$ such that $X$ and $Y$ are comonotonic (see comonotonicity).
Additivity for comonotonic risks is desirable because if one adopts subadditivity as a general rule, then it is unreasonable to have $\rho(X+Y) < \rho(X) + \rho(Y)$ because neither risk is a hedge against the other, that is, they move together. If a premium principle is additive for comonotonic risks, then is it layer additive. Note that Property 11 implies Property 6, if $\rho$ additionally satisfies a continuity condition.

**12. Monotonicity**

If $X$ and $Y$ are two losses such that $X \geq Y$, then $\rho(X) \geq \rho(Y)$

**13. Preserves first stochastic dominance (FSD) ordering**

If $S_X(t) \leq S_Y(t)$ for all $t \geq 0$, then $\rho(X) \leq \rho(Y)$.

**14. Preserves stop-loss (SL)ordering**

If $E(X-d)_+ \leq E(Y-d)_+$ for all $d \geq 0$, then $\rho(X) \leq \rho(Y)$
Property 1, together with Property 12, imply Property 13. Also, if $\rho$ preserves SL ordering, then $\rho$ preserves FSD ordering because stop-loss ordering is weaker. These orderings are commonly used in actuarial science to order risks (partially) because they represent the common ordering of groups of decision makers. Finally, we present a technical property that is useful in characterizing certain premium principles.

**15. Continuity**

Let $X \in \mathcal{X}$; then, $\lim_{a \to 0^+} \rho(\max(X-a, 0)) = \rho(X)$, and $\lim_{a \to \infty} \rho(\min(X, a)) = \rho(X)$.

**Concept of Coherence**

Several authors have selected some of these conditions to form a set of requirements that any risk measure should satisfy. The first class of risk measures which was introduced by Artzner et al [3] is the coherent risk measures. And was constructed to possess all mathematical properties to properly reflect the economy. And hence it takes the second step within the risk measurement procedure into account. A risk measure is called coherent if it satisfies the following axioms.

**Definition 2.5** (Coherence)**.**
*A risk measure $\rho$ that is translative, positive homogeneous, subadditive and monotone is called coherent.*

While the coherent risk measure provides a standard for constructing meaningful measure of risks, it is not sufficiently restrictive to specify a unique risk measure. Instead, it characterizes a large class of risk measure. Otherwise, It is worth mentioning that coherence is defined with respect to a set of axioms, and no set is universally accepted. Modifying the set of axioms regarded as desirable leads to other 'coherent' risk measures.

### 2.5.3 Some premium principles

The following premium principles are frequently encountered. For more details, we refer to Young (2004) [122] and Bühlmann (1970) [18].

#### a. Net Premium Principle

This premium principle does not load for risk. It is the first premium principle that many actuaries learn. It is widely applied in the literature because actuaries often assume that risk is essentially non-existent if the insurer sells enough identically distributed and independent policies.

$$\rho(X) = \mathbb{E}(X)$$

This premium also known as the equivalence principle; it is sufficient for a risk neutral insurer only.

#### b. Expected Value Premium Principle (level $\lambda$)

This premium principle builds on principle a, the Net Premium Principle, by including a proportional risk load. It is almost always used in life insurance and in risk theory. This principle is easy to understand and to explain to policyholders.

$$\rho(X) = (1+\lambda)\,\mathbb{E}(X)\,,\ \ \lambda > 0.$$

#### c. Variance Premium Principle (level $\beta$)

The premium principle also builds on the Net Premium Principle by including a risk load that is proportional to the variance of the risk. Bühlmann (1970) [18, chapter 4] studied this premium principle in detail.

$$\rho(X) = \mathbb{E}(X) + \beta Var(X)\,,\ \ \beta > 0.$$

**d. Standard Deviation Premium Principle (level $\alpha$)**

$$\rho(X) = \mathbb{E}(X) + \alpha\sqrt{Var(X)}, \ \alpha > 0.$$

The loading is proportional to the standard deviation of $X$. The loss can be written as

$$\rho(X) - X = \sqrt{Var(X)}\left(\alpha - \frac{X - \mathbb{E}(X)}{\sqrt{Var(X)}}\right).$$

or the loss is equal to the loading parameter minus a rv with mean value 0 and variance 1.

**e. Exponential Premium Principle**

This premium principle arises from the principle of equivalent utility when the utility function is exponential.

$$\rho(X) = \frac{1}{\theta}\log(\mathbb{E}(\exp(\theta X))), \text{ for some } \theta > 0.$$

**f. Esscher Premium Principle**

$$\rho(X) = \frac{\mathbb{E}(X\exp(\theta Z))}{\mathbb{E}(\exp(\theta Z))},$$

for some $\theta > 0$ and for some rv $Z$. Bühlmann (1980) [19] derived this premium principle when he studied risk exchanges.

**g. Principle of Equivalent Utility**

$\rho(X)$ solves the equation

$$u(w) = \mathbb{E}(u(w - X + \rho)),$$

where $u$ is an increasing, concave utility of wealth (of the insurer), and $w$ is the initial wealth (of the insurer).

**h. Proportional Hazards Premium Principle**

$$\rho(X) = \int_0^\infty (S_X(x))^\rho \, dx,$$

where $\rho > 0$ is called a risk index or distortion parameter. This parameter controls the amount of the risk loading included in the premium for given riskiness of the loss variable $X$. Wang (1996) [118] studied the many nice properties of this premium principle.

**i. Distortion Risk Premium Principle**

$$\rho(X) = \int_0^{\infty} g(S_X(x))\, dx,$$

where $g$ is an increasing function that maps $[0;1]$ into $[0;1]$. The function $g$ is called a distortion and $g(SX(x))$ is called a distorted (tail) probability.

Distortion risk premium principle have their origin in Yaari's (1987) [121] dual theory of choice under risk that consists in measuring the risks by applying a distortion function $g$ on the df $F$. The net premium principle and proportional hazards premium principle are a special case of distortion risk premium principle with the distortions $g$ given by $g(s) = s$ and $g(s) = s^{\rho}$ (resp.). See Wang (1996) [118] for other distortions.

### 2.5.4 Risk Measures

The essential technical tools to quantify risks are risk measures. Some of the most frequent questions concerning risk management in application involve extreme quantile estimation. This corresponds to the determination of the value a given variable exceeds with a given (low) probability. A typical example of such tail related risk measures is the Value-at-Risk (VaR) calculation. Other less frequently used measures are the expected shortfall and the return level.

**Value-at-Risk**

The last decade has seen a growing interest in quantiles of probability distributions on the part of practitioners. Since quantiles have a simple interpretation in terms of over- or undershoot probabilities they have found their way into current risk management practice in the form of the concept of value-at-risk abbreviated VaR. This concept was introduced to answer the following question: how much can we expect to lose in one day, week, year, ... with a given probability? In today's financial world, VaR has become the benchmark risk measure: its importance is unquestioned since regulators accept this model as the basis for setting capital requirements for market risk exposure.

VaR risk measure was developed in the 1990's as a response to financial disasters. Although developed in the 1990's, the methodology behind VaR is not new, it can be traced back to 1952 to the basic mean-variance framework of Markowitz. Moreover, the VaR principle was used in actuarial sciences long before it was reinvented for investment banking. Although,within actuarial sciences the more common phrase was the quantile risk measure as opposed to Value-at-Risk. This risk measure has the advantage of being relatively easy to evaluate and to understand. This made it very popular from the practitioner point. Informally, VaR can be defined as the worst loss over a target horizon such that with a prespecified probability that the actual loss will be higher. The formal mathematical definition is the following :

**Definition 2.6** (Value-at-Risk).
*Given a risk $X$ and a probability level $p \in (0,1)$, the corresponding VaR is a high quantile of the distribution of risk, typically the $95th$ or $99th$ percentile. That is*

$$VaR_p := F_x^{\leftarrow}(p), \tag{2.118}$$

*where $F^{\leftarrow}$ is the generalized inverse of the df $F$ of a certain risk $X$.*

It is worth mentioning that VaR's always exist and are expressed in the proper unit of measure, namely in lost money. Since VaR is defined with the help of the quantile function $F$, all their properties immediately apply to VaR. We will often resort to the following equivalence relation, which holds for all

$$VaR_p \leq x \Longleftrightarrow p \leq F_x(x) \tag{2.119}$$

VaR fails to be subadditive (except in some very special cases, such as when the $X_i$ are multivariate normal). Thus, in general, VaR has the surprising property that the VaR of a sum may be higher than the sum of the VaR's. In such a case, diversification will lead to more risk being reported. Consider two independent Pareto risks of parameter 1; $X$ and $Y$: Show that the inequality

$$VaR_p(X) + VaR_p(Y) < VaR_p(X+Y) \tag{2.120}$$

holds for any $p$; so that VaR cannot be subadditive in this simple case. A possible harmful aspect of the lack of subadditivity is that a decentralized risk management system may fail because VaR's calculated for individual portfolios may not be summed to produce an upper bound for the VaR of the combined portfolio.

### Tail Value-at-Risk

A single VaR at a predetermined level $p$ does not give any information about the thickness of the upper tail of the distribution function. This is a considerable shortcoming since in practice a regulator is not only concerned with the frequency of default, but also with the severity of default. Also shareholders and management should be concerned with the question 'how bad is bad?' when they want to evaluate the risks at hand in a consistent way. Therefore, one often uses another risk measure, which is called the tail value-at-risk (TVaR) and defined next.

**Definition 2.7** (Tail Value-at-Risk).

Given a risk $X$ and a probability level $p$, the corresponding TVaR, is defined as

$$TVaR_p = \frac{1}{1-p}\int_p^1 VaR_\xi d\xi, \ 0 < p < 1. \tag{2.121}$$

We thus see that $TVaR_p$ can be viewed as the 'arithmetic average' of the VaR's of $X$, from $p$ on.

### Conditional Tail Expectation

The conditional tail expectation (CTE) represents the conditional expected loss given that the loss exceeds its VaR.

**Definition 2.8** (Conditional Tail Expectation).
*For a risk $X$, the Conditional Tail Expectation (CTE) at probability level $p \in (0,1)$ is defined as*

$$CTE_p = \mathbb{E}\left(X \,|X > VaR_p\right). \tag{2.122}$$

So the CTE is the 'average loss in the worst $100(1-p)\%$ cases'. Writing $d = VaR_p$ we have a critical loss threshold corresponding to some confidence level $p$, $CTE_p$ provides a cushion against the mean
value of losses exceeding the critical threshold $d$.

#### Conditional VaR

An alternative to CTE is the conditional VaR (or CVaR). The CVaR is the expected value of the losses exceeding VaR.

$$\begin{aligned} CVaR_p &= \mathbb{E}\left(X - VaR_p \,|X > VaR_p\right) \\ &= CTE_p - VaR_p \end{aligned} \tag{2.123}$$

It is easy to see that CVaR is related to the mean-excess function through

$$CVaR_p = e_X\left(VaR_p\right). \tag{2.124}$$

Therefore, evaluating the mef at quantiles yields CVaR.

#### Expected Shortfall

Artzner et al.(1997 [3],1999 [4]) show that the VaR has various theoretical deficiencies as a measure of risk. They conclude that the VaR is not a coherent measure of risk as it fails to be subadditive in general. On the other hand, VaR gives only a lower limit of the losses that occur with a given frequency, but tell us nothing about the potential size of the loss given that a loss exceeding this lower bound has occurred. These authors propose the use of the so-called expected shortfall or tail conditional expectation instead. The expected shortfall measures the expected loss given that the loss $l$ exceeds VaR. In particular, this risk measure gives some information about the size of the potential losses given that a loss bigger than VaR has occurred. Expected shortfall is a coherent measure of risk as defined by Artzner et al. (1999) [4]. Commonly speaking, the ES addresses the important question: "*given that we will have a bad day, how bad do we expect it to be*"?. Formally, the expected shortfall for risk $X$ and high confidence level $p$ is defined as follows:

**Definition 2.9** (Expected shortfall)**.**
*The ES of a loss $X$ is the expected value of the losses in excess of the VaR. That is*

$$ES_p := E\left(\left(X - VaR_p\right)_+\right). \tag{2.125}$$

*where $p$ is as in definition 2.6.*
*The ES is the stop-loss premium with retention $VaR_p$.*

#### Return Level

**Definition 2.10** (Return level).
*If $H_\theta$ is the distribution of the maximum observed over successive non overlapping periods of equal length, the return level*

$$R_m = R_m(l) := H_\theta^{\leftarrow}(1 - 1/m), m \geq 1, \tag{2.126}$$

*is the expected level to be exceeded in one out of $m$ periods of length $l$.*

The return level can be used as a measure of the maximum loss of a portfolio, a rather more conservative measure than the Value-at-Risk.

### 2.5.5 Relationships Between Risk Measures

The following relation holds between the first three risk measures defined above.

**Proposition 2.3.**
*For any $p \in (0,1)$, the following identities are valid :*

$$TVaR_p = VaR_p + \frac{1}{1-p} ES_p \tag{2.127}$$

$$CTE_p = VaR_p + \frac{1}{\bar{F}_X(VaR_p)} ES_p \tag{2.128}$$

$$CVaR_p = \frac{ES_p}{\bar{F}_X(VaR_p)}. \tag{2.129}$$

**Proof.** See Denuit et al [38]. □

**Corollary 2.3.**
*Note that if $F_X$ is continuous then by combining (2.127) and (2.128) we find*

$$CTE_p = TVaR_p, \ p \in (0,1), \tag{2.130}$$

*so that CTE and TVaR coincide for all p in this special case. In general, however, we only have*

$$TVaR_p = CTE_p + \left( \frac{1}{1-p} - \frac{1}{\bar{F}_X(VaR_p)} \right) ES_p. \tag{2.131}$$

*Since the quantity between the brackets can be different from 0 for some values of p, TVaR and CTE are not always equal.*

### 2.5.6 Estimating Risk Measures

In 2003, empirical estimation of risk measures and relative quantities have proposed by Jones and Zitikis [74]. Kaiser and Brazauskas [76] proposed the confidence interval estimation of various risk measures in the case where the variance is finite as : Proportional Hazards Transform (PHT), Wang Transform (WT), Value-at-Risk (VaR), and Conditional Tail Expectation (CTE).

Since the risk measures above are actually high quantiles (or function of a high quantile for the ES), their estimations are straightforward applications of the results of Section 2.4. Consequently, all the properties of extreme quantile estimators are inherited.

Any quantile estimator seen in Section 2.4 can be used to estimate the VaR with, however, a preference for the POT based estimator of relation.

**Proposition 2.4** (Estimating $\mathrm{VaR}_p$).
*For $n \geq 1$, let $(X_1, \ldots, X_n)$ be a sample from a loss $X$. If $u$ is a fixed threshold and $N_u$ the number of observations exceeding $u$, then $VaR_p$ is estimated by*

$$\widehat{VaR_p} := u + \frac{\hat{\sigma}_u}{\hat{\gamma}_u}\left(\left(\frac{N_u}{np}\right)^{\hat{\gamma}_u} - 1\right), 0 < p < 1, \tag{2.132}$$

where $\hat{\gamma}_u$ and $\hat{\sigma}_u$ are the estimates of the parameters of the fitted GPD.

For the estimation of the ES, notice that it is related to the VaR by

$$ES_p = VaR_p + \mathbb{E}\left(X - VaR_p \,|X > VaR_p\right). \tag{2.133}$$

If $Y := X - u$ denotes the excess over threshold $u$, then

$$ES_p = VaR_p + \mathbb{E}\left(Y - z\,|Y > z\right), \tag{2.134}$$

for $z := VaR_p - u$. The second term of the right hand side is the mef of $Y$ over threshold $z$. Assuming that $Y$ has a GPD with parameters $\gamma < 1$ and $\sigma > 0$, then by property (a) of Proposition 1.18 we have

$$\mathbb{E}\left(Y - z\,|Y > z\right) = \frac{\sigma + \gamma\left(z\right)}{1 - \gamma}, \sigma + \gamma\left(z\right) > 0. \tag{2.135}$$

This leads to

$$ES_p = VaR_p + \frac{\sigma + \gamma\left(VaR_p - u\right)}{1 - \gamma}. \tag{2.136}$$

Substituting $\widehat{VaR}_p$, $\hat{\gamma}_u$ and $\hat{\gamma}_u$ and $\hat{\sigma}_u$ for $VaR_p$, $\gamma$ and $\sigma$ respectively, yields the following estimate for the ES.

**Definition 2.11** (Mean-excess function)**.**
*Given a non-negative rv $X$, the associated mean-excess function (mef) is defined as*

$$e_X(x) = E\left[X - x \,|X > x\right], x > 0 \tag{2.137}$$

**Proposition 2.5** (Estimating $\mathrm{ES}_p$)**.**

$$\widehat{ES}_p := \frac{\widehat{VaR}_p}{1 - \hat{\gamma}_u} + \frac{\hat{\sigma}_u - \hat{\gamma}_u}{1 - \hat{\gamma}_u} \tag{2.138}$$

Finally, to estimate the return level, it suffices to substitute $1/m$ for $p$ in relation (2.90).

**Proposition 2.6.**

$$\hat{R}_m := \begin{cases} \hat{\mu} - \dfrac{\hat{\sigma}}{\hat{\gamma}}\left(1 - \left(-\log\left(1 - 1/m\right)\right)^{-\hat{\gamma}}\right) & \text{if} \quad \gamma \neq 0, \\ \hat{\mu} - \hat{\sigma} \log\left(-\log\left(1 - 1/m\right)\right) & \text{if} \quad \gamma = 0, \end{cases} \tag{2.139}$$

*where $\hat{\gamma}$, $\hat{\mu}$ and $\hat{\sigma}$ are estimates of the parameters of the parameters of the GEV distribution $H_\theta$.*

# Part II

# Main Results

# Chapter 3

# Complete Flood Frequency Analysis in Abiod Watershed Biskra (Algeria)

## Contents

Extreme hydrological events, such as floods and droughts, are one of the natural disasters that occur in several parts of the world. They are regarded as being the most costly natural risks in terms of the disastrous consequences in human lives and in property damages. In this chapter we will estimate flood events of Abiod wadi at given return periods at the gauge station of M'chouneche, located closely to the city of Biskra in a semiarid region of southern east of Algeria. This is a problematic issue in several ways, because of the existence of a dam to the downstream, including the field of the sedimentation and the water leaks through the dam during floods. The considered data series is new. A complete frequency analysis (FA) is performed on a series of observed daily average discharges, including classical statistical tools as well as recent techniques. It is noteworthy that the content of this chapter consists in the work that I jointly made with Professors Meraghni, Benkhaled, Chebana and Necir, and which was published in 2016 in Natural Hazards journal [10].

## 3.1 Study Area and Data

In this section, we present the region where the site of interest is located, followed by a description of the available data.

### 3.1.1 Study Area

The Abiod wadi watershed, with an area of 1300 $km^2$, is located in the Aurès massif in the southern east of Algeria in North Africa (Figure 3.1). It is part of the endorheic watershed Chott Melghir. The wadi length is 85 $km$ from its origin in the Chelia (2326 $m$ high) and Ichemoul (2100 $m$ high) mountains. After crossing Tighanimine, the wadi gradually flows into the canyons of Ghoufi and M'chouneche gorges and then opens a path to the plain until the Saharian gorge Foum El Gherza. The valley of the wadi is mainly composed of sedimentary rocks, comprising alternating limestone, marl, soft sediments (sandstones, conglomerates) and some evaporates (gypsum) dated of Paleogene.

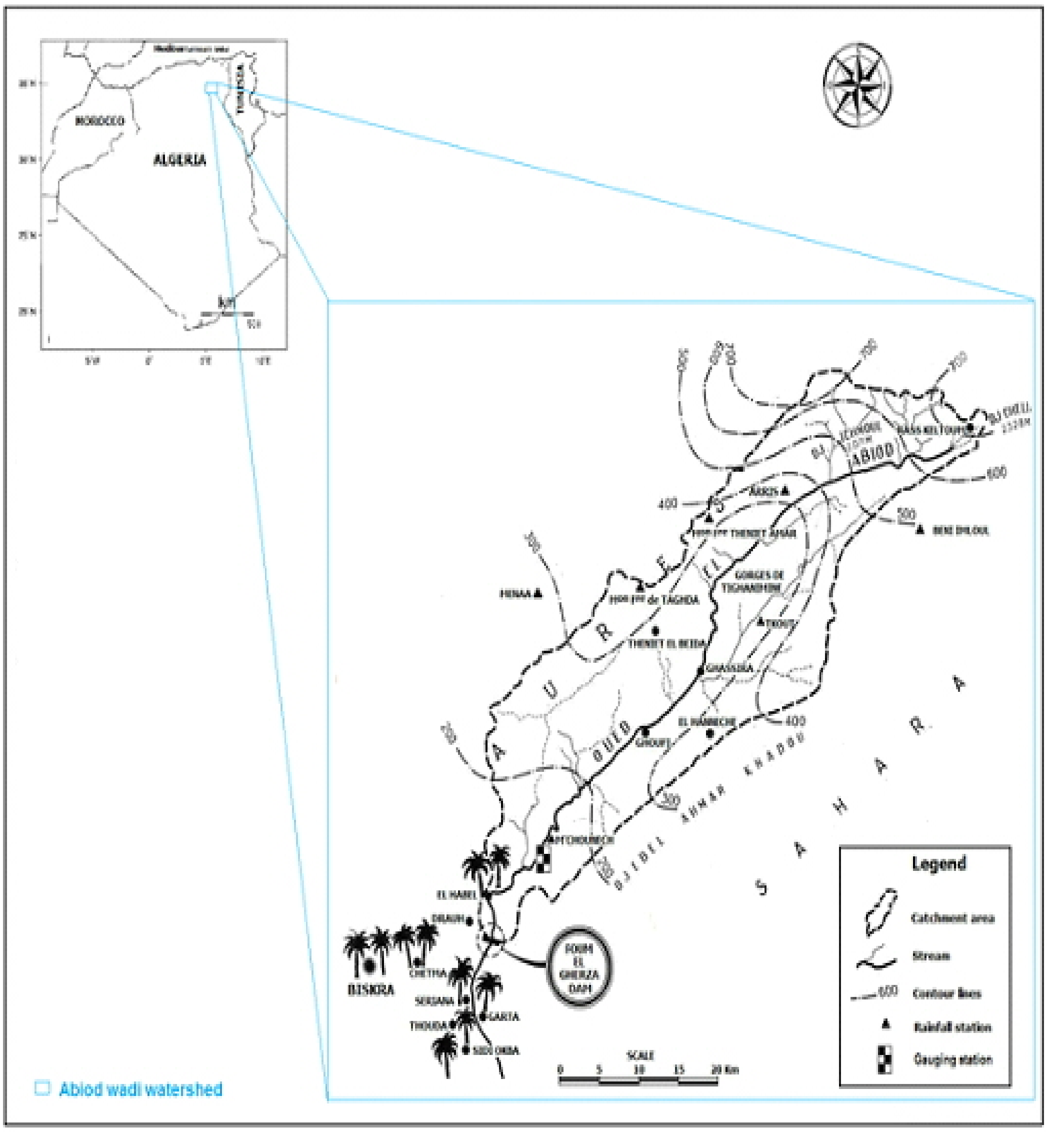



FIGURE 3.1. Geographical location of the Abiod wadi watershed

The watershed is characterized by its asymmetry, a mountainous area in the north to over 2000 $m$ (Chelia) and another low area in the south (El Habel 295 $m$). The relief is rugged with slopes ranging between 12.5 and 25% for half of the area, and from 3 to 12.5% for another 40% of the area. Land cover is a mix of rocky outcrops, highly eroded soil, sparse vegetation, a few forests, crops, gardens and pastures (Hamel 2009) [63]. In the orographic and hydrographic points of view, Abiod wadi is characterized by two distinct climatic regions: the Aurès, where rainfall averages 450 $mm/year$, and the Sahara plain with mean rainfall $100 - 150$ $mm/year$. The climate of Abiod wadi watershed is thus semiarid to arid. Along Abiod wadi to the Foum El Gherza dam, there are six rainfall stations, and one hydrometric station is located 18 km upstream of the dam, as shown in Figure 3.1, which was damaged during the floods of 1994–1995 and it is not operational since.

The choice of this station was made on the basis of climatic context of the study area. It is the only station on the studied basin, and it is rather representative of the whole southeast region in Algeria, which is arid to semiarid. Also, the size of the series used shows the interest of the FA application.

### 3.1.2 Data Description

The data set used in this study is provided by the National Agency of Hydraulics Resources (ANRH) of Biskra, and it is the first time to be considered and studied. It consists of the daily average discharges $Q_1, \ldots, Q_N$ (with $N = 8034$), collected at the gauge station of M'chouneche over 22 years from 1972 to 1994.

Note that the IACWD Bulletin 17B (1982) [70] suggests that at least 10 years of record is necessary to warrant a statistical analysis. For instance, Tramblay et al. (2008) [115] used a minimum of 10 years of daily data. The short data size can affect the choice of distributions, the quantile estimations, particularly those corresponding to large return periods and the extent of confidence intervals. The size of the used data in the present study is relatively large, to perform a frequency analysis (FA), as in a number of similar studies (Chebana et al. 2009) [23].

## 3.2 Methodology

In this section, after defining the type of series to be analyzed, namely the POT series, we briefly present the required elements to perform a hydrological FA. The latter is a statistical approach of prediction commonly used in hydrology to relate the magnitude of extreme events to a probability of their occurrence (Chow et al. 1988) [26]. It allows, for the selected station, to estimate the flood quantiles of given return periods. In general, FA involves four main steps

1. characterization of the data and determination of the usual statistical indicators, such as the mean, the standard deviation (SD), the coefficients of skewness (Cs), kurtosis (Ck) and variation (Cv) and detection of outliers,

2. checking the basic hypotheses of FA, i.e., homogeneity, stationarity and independence, applicability on the studied data set,

3. fitting of probability distributions, estimation of the associated parameters and selection of the best model to represent the data and

4. risk assessment based on quantiles or return periods (e.g., Bobée and Ashkar 1991 [16]; Chebana and Ouarda 2011 [22]; Haktanir 1992 [59]; Rao and Hamed 2000 [100]).

### 3.2.1 Peaks Over Threshold Series

The data to be extracted and then used in this approach consist in the observations that exceed a selected relatively high threshold $u$. Let $Q$ represent the daily average discharge and denoted by $N_u$ is the number of discharges exceeding $u$. Then, the sample of excesses is defined as

$$\{E_j := Q_{ij} - u \ \ s.t. \ Q_{ij} > u \ ; \ j = 1, \dots N_u\}. \tag{3.1}$$

In this approach, the selection of an appropriate threshold is crucial. This approach is useful and has some advantages compared to the AM one, even though the latter is widely used. It is of particular interest in situations where the AM could not perform well especially in situation with little extreme data or the extracted extremes by AM cannot be considered as extremes in a physical or hydrological meaning.

### GPD Approximation

Statistically, the distribution of the POT series $E_1, \ldots, E_{Nu}$ can be determined by making use of the GPD which is a cdf $G_{\gamma,\sigma}$ defined, for $x \in S(\gamma,\sigma) := [0,\infty)$ if $\gamma \geq 0$ and $[0, -\sigma/\gamma)$ if $\gamma < 0$, by

$$G_{\gamma,\sigma}(x) := \begin{cases} 1 - \left(1 + \gamma \dfrac{x}{\sigma}\right)^{\frac{-1}{\gamma}} & if \quad \gamma \neq 0, \\ 1 - \exp\left(-x/\sigma\right) & if \quad \gamma = 0, \end{cases} \tag{3.2}$$

where $\gamma \in \mathbb{R}$ and $\sigma > 0$ are, respectively, shape and scale parameters (Hosking and Wallis 1997) [68].

Let $Fu(x) := P(Q - u \leq x | Q > u)$ denote the excess cdf of $Q$ over a given threshold $u$ . Then, we have the following result

$$\lim_{u \longrightarrow q_F} \sup_{0<x<q_F-u} \left| F_u\left(x\right) - G_{\gamma,\sigma(u)}\left(x\right) \right| = 0, \tag{3.3}$$

where $q_F$ is the right end point of the cdf $F$. This result, due to Balkema and de Haan (1974) [6] and Pickands (1975) [97], is one of the most useful concepts in statistical methods for extremes. It says that for large threshold $u$ , the excess cdf $F_u$ is likely to be well approximated by a GPD.

### Threshold Selection

In order to obtain the asymptotic result in (3.3), the threshold $u$ should be large enough which has as a consequence a satisfactory GPD approximation. The choice of the threshold is a crucial issue in the POT procedure. Indeed, selecting a threshold that is too low results in a large bias in the estimation, whereas taking one that is too high yields a big variance (Embrechts et al. 1997, Sects. 6.4 and 6.5) [47]. Hence, a compromise between bias and variance is to be found. To this end, one can minimize the asymptotic mean squared error, which is composed by the bias and variance. Furthermore, several graphical procedures are available to select $u$, such as the mean residual life (MRL), threshold choice (TC) and dispersion index (DI) plots. On the other hand, the choice of $u$ can be based on physical considerations, e.g., by identifying the flood level of the river of interest. For a survey of the main selection procedures, see, e.g., the paper of Lang et al. (1999) [84].

### 3.2.2 Exploratory Data Analysis

The first step allows to check the data quality and to screen the data to avoid outlier effects. It also permits to obtain prior information, e.g., the shape, regarding the distribution to be selected. The presence of outliers in the data can have an important effect and causes difficulties when fitting a distribution (Ashkar and Ouarda 1993) [5] especially on the distribution upper part. The Grubbs and Beck (1972) [53] statistical test, based on the assumption of normality data, is designed to detect low and high outliers. In the case where the original data are not normal, they should be appropriately transformed. According to Section 1.8.3 in Rao and Hamed (2000) [100], this test is based on the following quantities

$$x_H := \exp\left(\bar{x} + k_n s\right), \tag{3.4}$$

$$x_L := \exp\left(\bar{x} - k_n s\right), \tag{3.5}$$

where $\bar{x}$ and $s$ are, respectively, the mean and standard deviation of the natural logarithms of the sample, and $k_n$ is the Grubbs–Beck statistic tabulated for various sample sizes and significance levels by Grubbs and Beck (1972) [53]. For instance, at the 10% significance level, the following approximation is used

$$k_n := -3.62201 + 6.28446n^{1/4} - 2.49835n^{1/2} + 0.491436n^{3/4} - 0.037911n, \tag{3.6}$$

where $n$ is the sample size.

The observations greater than $x_H$ are considered to be high outliers, while those less than $x_L$ are taken as low outliers.

### 3.2.3 Testing Independence, Stationarity and Homogeneity

Three basic assumptions are required to correctly apply FA of extreme hydrological events, namely independence, stationarity and homogeneity of the data (Bobée and Ashkar 1991) [16]. To verify these assumptions, three tests are widely used in the literature. The Wald-Wolfowitz test is employed for the independence, the homogeneity test of Wilcoxon is applied to check whether the data come from the same distribution or not, and the Mann-Kendall test allows to verify stationarity of the data, i.e., the series does not present a trend over time. These three tests have the advantage of being nonparametric and are widely used in hydrological FA. In other words, they do not require any prior knowledge on the distribution of the data.

### 3.2.4 Parameter Estimation and Model Selection

The choice of the appropriate model is one of the most important issues in FA. In practice, the distribution of hydroclimatic series is not known. Using the fitted probability distribution, it is possible to predict the probability of exceedance for a specified magnitude, i.e., quantile, or the magnitude associated with a specific exceedance probability. To estimate the parameters associated with the appropriate probability distribution, popular techniques are used in hydrology, including the methods of maximum likelihood (ML) (e.g., Clarke 1994 [27]; Natural Environment Research Council 1975 [92]), moments (MM) and probability weighted moments (PWM) (e.g., Chebana et al. 2010 [24]; Hosking et al. 1985 [69]). The latter is equivalent to the L-moment method which is widely used in hydrological FA (Hosking 1990) [66].

The choice of the adequate distribution is determined on the basis of numerous classical and recent statistical tools, including graphical representations (Institute of Hydrology 1999 [72]; Natural Environment Research Council 1975 [92]) and goodness-of-fit tests such as the tests of Pearson (Chi-squared, Chi2), Kolmogorov-Smirnov (KS), Cramer-von Mises and the normality-specific Shapiro-Wilk (SW) test. Due to the importance of the distribution impact in FA, these tools should be exploited. This point is widely studied in the literature (Benkhaled et al. 2014 [12]; Ehsanzadeh et al. 2010 [43]; El Adlouni et al. 2008 [45]; Hebal and Remini 2011 [64]; Hosking and Wallis 1997 [68]; Koutsoyiannis 2003 [81]; Ouarda et al. 1994 [96]).

Nonetheless, the decision procedures mentioned above are not perfectly suited for extreme value distributions, because they are not sensitive enough to deviations in the tails. Several transformations have been proposed to overcome the limitations of the aforementioned tests (Khamis 1997 [78]; Laio 2004 [82]; Liao and Shimokawa 1999 [85]). In our application, where we focus on the upper tail of the distribution, we perform the Anderson-Darling k-sample test ($k = 2$) implemented in the adk package of the statistical software R. This procedure is used to test the null hypothesis that $k$ samples come from one common continuous distribution. In our case, the first sample of size 42 is the considered POT series and the second one consists in values generated from the GPD model. For more details on this test, we refer to Scholz and Stephens (1987) [107].

The probability distributions that are appropriate for hydrology data are those with heavy tails. A number of them are listed, e.g., in Kite (1988) [79], Rao and Hamed (2000) [100] and Salas and Smith (1980) [105]. In order to select the appropriate distribution among those which passed the goodness-of-fit tests, one or more criteria are required. To this end, one can consider the Akaike and Bayesian information criterion (AIC, BIC), respectively, proposed by Akaike (1974) [1] and Schwartz (1978) [108]. They are given by

$$AIC := -2\ln L + 2k, \tag{3.7}$$

$$BIC := -2\ln L + 2k\ln m, \tag{3.8}$$

where $L$ is the likelihood function, $k$ the number of parameters and $m$ the sample size. The best fit is the one associated with the smallest criterion AIC or BIC values (Ehsanzadeh et al. 2010 [43]; Hebal and Remini 2011 [64]; Rao and Hamed 2000 [100]).

### 3.2.5 Quantile Estimation

Once the appropriate distribution is selected, the quantiles and return periods can be evaluated. The quantile estimation for various recurrence intervals is the main goal in hydrological practice. The notion of return period for hydrological extreme events is commonly used in FA, where the objective is to obtain reliable estimates of the quantiles corresponding to given return periods of scientific relevance or government standard requirements (Rao and Hamed 2000) [100]. In the FA context, the uncertainty decreases with the sample size, whereas it increases with the return period when estimating quantiles.

In many environmental applications, the sample size is rarely sufficient to enable good extreme quantiles estimations. Usually, a quantile of return period $T$ can be reliably estimated from a data record of length $n$ if $T < n$. However, in many cases, this condition is rarely satisfied-since typically $n < 50$ for hydrological applications based on annual data (Hosking and Wallis 1997) [68].

## 3.3 Results and Discussion

The application of the presented methodology in Section 3.2 to the data described in Section 3.1 leads to the following results, obtained by means of the packages stats, evir and POT of the statistical software R (Ihaka and Gentleman 1996) [71] and also by using the HYFRAN-PLUS software, El Adlouni and Bobée (2010) [44].

### 3.3.1 Exploratory Analysis and Outlier Detection

From Figure 3.2, it appears that the whole daily data series varies from a minimum value of 0 $m^3/s$ corresponding to many dry days to a maximum value of 78.57 $m^3/s$ recorded on September 21, 1989. The average flow of 0.39 $m^3/s$ is a relatively low in comparison with other tributary wadis of Chott Melghir like El Hai wadi and Djamorrah wadi (Mebarki 2005) [88]. The standard deviation of 2.48 $m^3/s$ yields a coefficient of variation equal to 6.39.

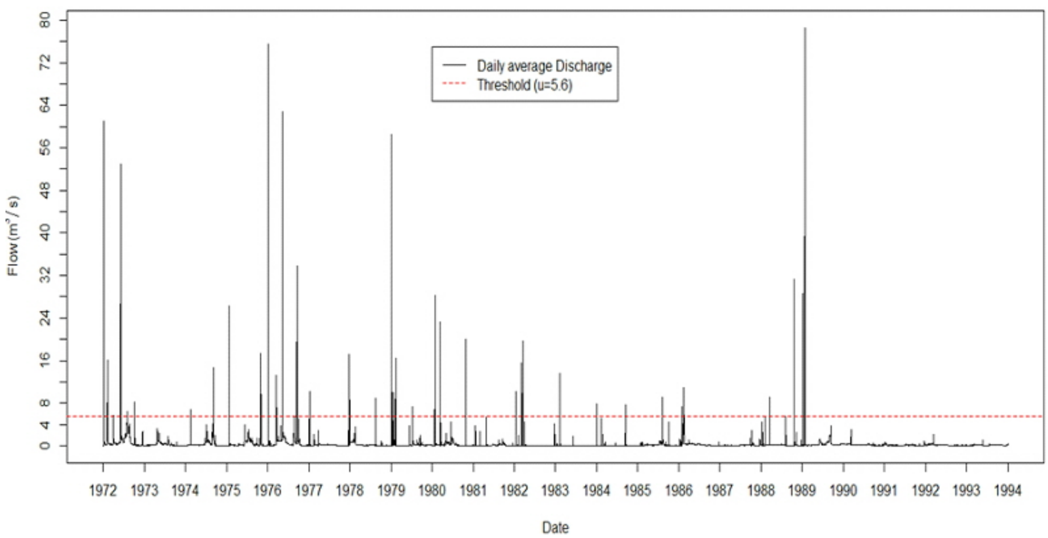



FIGURE 3.2. Time series plot of the daily average discharge at M'chouneche station covering the period 01/09/1972–31/08/1994

The boxplot in Figure 3.3 clearly shows the existence of extreme values. Indeed, the median (0.10 $m^3/s$) is close to both $25th$ and $75th$ percentiles (0.04 and 0.20 $m^3/s$ ). In addition to this graphical consideration, the values of skewness and kurtosis (20.51 and 498.59 $m^3/s$ , respectively) eliminate the Gaussian model. In particular, the very large value of the kurtosis indicates longer and fatter distribution tails, urging us to focus on heavy-tailed models.

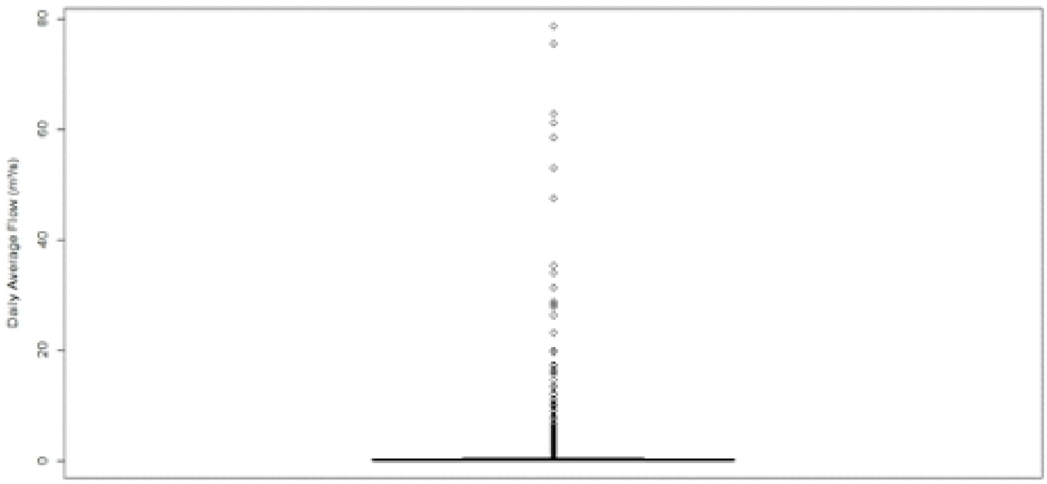



Figure 3.3. Boxplot of daily average discharge at M'chouneche station

From Figure 3.2, we observe high inter-annual and the short sample size (resulting from selection AM) which leads to selecting low discharges during the driest years, whereas some interesting discharges were not selected during the years where several floods have occurred. This explains the non relevance of the AM approach for Abiod wadi data analysis and suggests that the POT approach would be more appropriate and would lead to a more homogeneous sample of extreme discharges. This method starts with the selection of a convenient threshold and then the consideration of the observations that exceed this threshold.

In order to detect outliers, the quantities $x_H$ and $x_L$ are found to be 508.31 and 0.08, respectively. Since there is no value greater than $x_H$ and nor less than $x_L$, we conclude that, at the significant level of 10%, no outlier exist among the excesses. Since it is difficult to use the outlier detection test with the analysis of extremes and due to the lack of regional weather data, the significance level to 10% is considered.

### Threshold Selection

In this study, we adopt one of the available graphical tools, namely the TC-plot. From Figure 3.4, we can choose a threshold value $u = 5.6\ m^3/s$, which results in an excess series of size 58. However, as recommended by many authors (Beran and Nozdryn-Plotnicki 1977 [13]; Lang et al. 1999 [84]; Todorovic and Zelenhasic 1970 [114]), this data set must be reduced in order to avoid the effects of dependence. We eliminated the peaks being obviously part of the same flood, and in order to keep the character of flood seasonality, we retain three peaks per year over the recorded period. Thus, the length of the data series becomes 42. Figure 3.5 shows the distribution of these excesses, and Table 3.1 summarizes their elementary statistics.

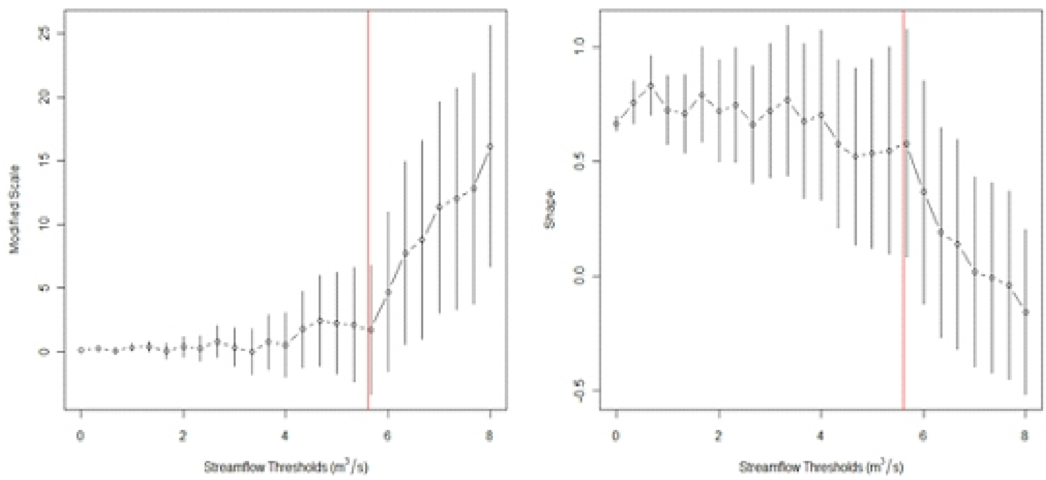
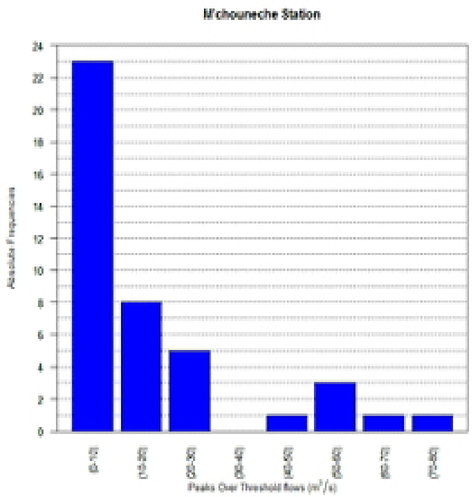



FIGURE 3.4. Graphical results of threshold selection applied for daily average discharge of Abiod wadi at M'chouneche station (TC-plot), vertical line corresponding to the threshold

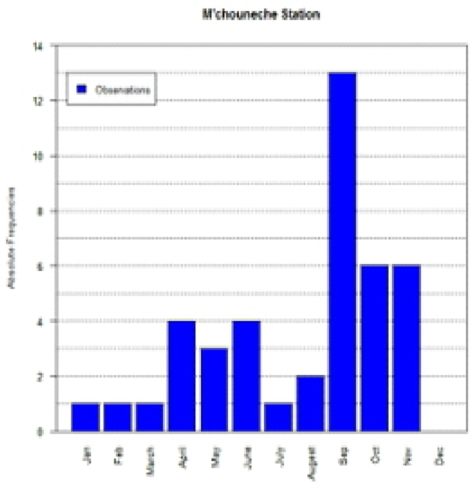
| | |
|---|---|
| Size | 42 Observations |
| Minimum | 0.02 ($m^3/s$) |
| $Qu_1$ (25$th$ percentile) | 3.36 ($m^3/s$) |
| Median | 7.83 ($m^3/s$) |
| Average | 15.72 ($m^3/s$) |
| $Qu_2$ (75$th$ percentile) | 19.92 ($m^3/s$) |
| SD | 19.70 ($m^3/s$) |
| Maximum | 72.97 ($m^3/s$) |
| Cs | 1.62 |
| Ck | 4.48 |

TABLE 3.1. Statistics summary of excess data set.

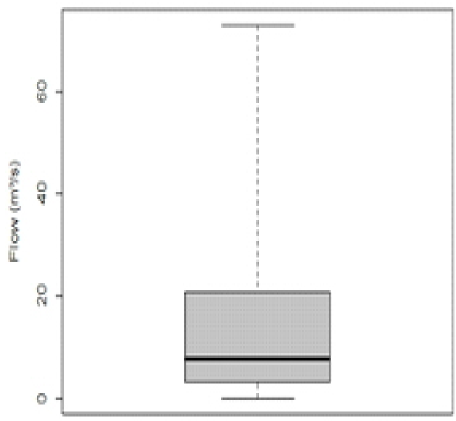
The positive skewness coefficient $Cs = 1.62$ reveals that the data are right-skewed relative to the mean excess, as shown in Figure. 3.5a. In Figure. 3.5a, the data are arranged by classes, of length 10 $m^3/s$ each, with the associated frequencies. It can be seen that some values are more frequent than others and that the majority of excesses have a low value varying between 0 and 10 $m^3/s$. Figure 3.5b, where the data are arranged according to the months of appearance, shows that the peaks generally occur in the fall season.




a)




b)




c)

FIGURE 3.5. Distribution of excess series at M'chouneche station a histogram by flow classes, b histogram by month and c boxplot

### 3.3.2 Testing the Basic FA Assumptions

The results of the required hypothesis testing on the considered data are given in Table 3.2. Applying Wilcoxon, Kendall and Wald-Wolfowitz tests, respectively, we conclude that the homogeneity, stationarity and independence of the excesses are accepted at any of the standard significance levels (1, 5 and 10%).Note that for the homogeneity test, we split the data in two sub-series 1972–1981 and 1982–1994 (any other subdivision led to the same conclusion). The homogeneity is also shown in Figure. 3.5a where there is only one mode (the highest frequency).

| Tests | Statistic value | p-values |
|---|---|---|
| Stationarity (Kendall) | 0.48 | 0.63 |
| Independence (Wald–Wolfowitz) | 0.94 | 0.35 |
| Homogeneity (Wilcoxon) | 0.79 | 0.43 |

Table 3.2. Stationarity, independence and homogeneity tests results.

### 3.3.3 Model Fitting

To fit a statistical distribution, we consider three commonly used estimation methods of the GPD parameters (ML, MM and PWM). Then, we perform the Anderson-Darling test to check the goodness of fit of the model. The results are summarized in Table 3.3. In view of the large p-values, we deduce that the GPD can be accepted as an appropriate model for the excess at any standard significance level (for instance 5%).

| Estimation method | Scale | Shape | Statistic value (Anderson–Darling) | p-values | AIC | BIC |
|---|---|---|---|---|---|---|
| ML | 10.19 | 0.39 | -0.55 | 0.49 | 315.68 | 326.63 |
| MM | 12.86 | 0.18 | -0.83 | 0.58 | 316.61 | 327.56 |
| PWM | 10.10 | 0.36 | -0.86 | 0.59 | 315.72 | 326.68 |

Table 3.3. GPD parameter estimation, Anderson–Darling goodness-of-fit test and information criterion results.

To discriminate between the obtained models, we use the AIC and BIC criteria. The last two columns of Table 3.3 as well as Figure 3.6b favor the GPD of the

ML fitting method. We illustrate the goodness of fit of the excesses to this model in Figure. 3.6a. Furthermore, this ML-based will be used for quantile estimation in the following section.

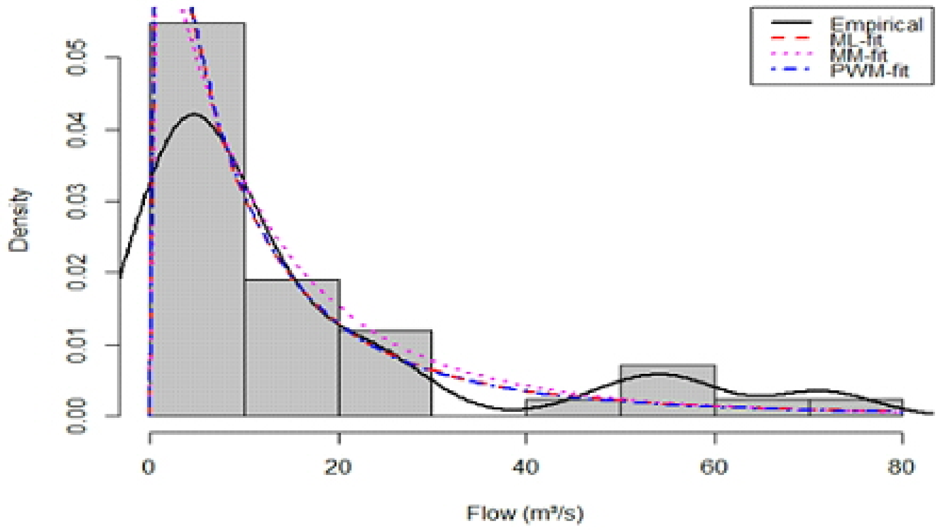
Note that the ML and PWM results are very similar, whereas those of the MM results are slightly different, but remain in the same range.

### 3.3.4 Quantile Estimation

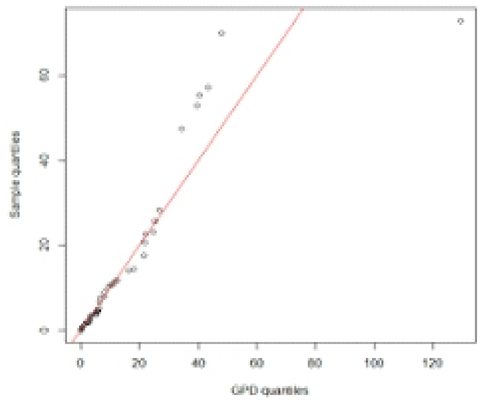
The estimation of extreme quantiles for different return periods should take into consideration the record period and the right tail of the distribution. The formally gauged record represents a relatively small sample of a much larger population of flood events. Thus, the extrapolation for long return periods is less accurate. In the M'chouneche station, only the following return periods were considered for the estimation of quantiles: 2, 5, 10, 20 and 50 years as presented in Table 3.4. The return period of the strongest stream flow in the 1972–1994 period, equal to 78.57 $m^3/s$, is estimated by means of Pareto's fitted model to be 30.62 years.

| Return period (years) | Estimated quantile ($m^3/s$) |
|---|---|
| 2 | 8.11 |
| 5 | 22.80 |
| 10 | 37.96 |
| 20 | 57.82 |
| 50 | 93.82 |

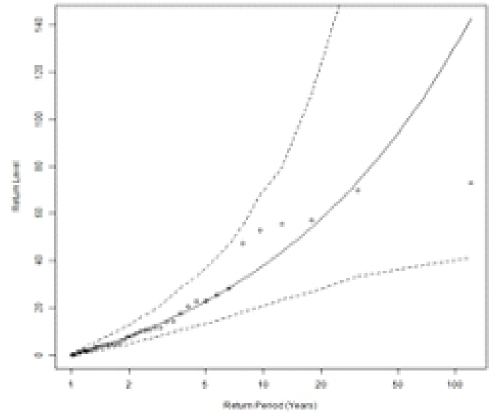
TABLE 3.4. Estimated quantiles of excess flows from the ML-based GPD.




a)



b)



c)

FIGURE 3.6. Best-fitted distributions of excess flows at M'chouneche station a distributions, b) qq plot of ML-based GPD and c) return level plot (95% confidence interval)

The confidence interval is a way to assess the uncertainty in the estimation of the distribution parameters and quantiles. For the GPD, the confidence bounds are obtained through asymptotic results (Hosking and Wallis 1997). In the present case study, one can see from Figure 3.6c that the GPD agrees with the observations for return levels less than 30 but not beyond even though they are all included in the confidence interval. This is probably due to the small number of peaks over the chosen threshold. Therefore, it is important to consider this distribution with care with return periods greater than 30 years. This point indicates the issue of the quantity of the required data in this station for better estimation of high return periods.

# Conclusion

The study of the Algerian wadis floods remains a quasi-unknown field as only some very specific indications are given in the Algerian hydrological directories. The present study is carried out in southern east of Algeria with new data series, in the context of FA. Mean daily discharges data recorded at the gauging station of M'chouneche in Abiod wadi, near Biskra, are available and considered in this study. Due to the high inter-annual variability of the data as well as to the relatively short record length, the AM approach is not adapted to this analysis. Hence, in this work, we considered a more appropriate procedure, namely the POT methodology.

Extreme values theory offers interesting conclusions when applied to the hydrological world. A presentation of this methodology has been made. The purpose of this thesis is to provide a suitable model for the excesses over a chosen threshold. This allows to estimate extreme flood events of given return periods. A complete FA was applied including appropriate tools, commonly used in hydrology. The issue of threshold selection was dealt by means of a graphical tool. Several fitting methods have led to different GPD models, and according to the results, the ML-based distribution was adopted. Because of the short record length, only return periods of 2, 5, 10, 20 and 50 years were considered. It was found that most of the extracted data corresponded to frequent events. In the present case study, the GPD distribution provided good estimates of return periods less than 30 years, but for higher values, the estimation is not acceptable and it is associated with high uncertainty (large confidence interval).

As a conclusion, we should emphasize that, in addition to the quality of data and sample size, the right GPD model heavily depends on the threshold which has to be very suitably chosen. To improve the flood FA at this site, future studies should focus on the importance of data monitoring. This issue is of primary importance as accurate data are very crucial and constitute the basis to any right conclusion that will be especially beneficial for local government and decision-makers.

Despite the weaknesses that it might have, this study has the merit of being the first of its kind to be performed in this area. It also opens interesting perspectives for future works and studies in the region, among which we can mention the regional frequency analysis and the multivariate frequency analysis. In the latter, one may apply the multivariate extreme value theory together with the copula tool to analyze the floods with respect to peaks, duration and volume simultaneously.

# Abbreviations and Notations

| **Abbreviations or Notations** | **Explanation** |
|---|---|
| $\lvert . \rvert$ | absolute value |
| AIC | Akaike Information Criterion |
| $\overset{a.s.}{\rightarrow}$ | almost safte convergence |
| AM | annual maximum |
| $\bar{X}_n$ | arithmitic mean |
| $S_n$ | arithmetic sum |
| BIC | Bayesian Information Criterion |
| $B\left(.,.\right)$ | beta function, defined by $B(p,q) := \int\limits_0^1 t^{p-1}\left(1-t\right)^{q-1} dt,\ p,q>0$, for $a,b>0$ |
| $B\left(.,.\right)$ | beta distribution |
| $Bin(n,p)$ | binomial distribution with succes probability $p$ |
| CLT | Central Limit Theorem |
| Chi2 | Chi-squared test |
| Cs | Coefficients of skewness |
| Ck | Coefficients of kurtosis |
| Cv | Coefficients of variation |
| $\binom{n}{i}$ | combination |
| $\overset{p}{\rightarrow}$ | convergence in probability |
| $\overset{d}{\rightarrow}$ | convergence in distribution |
| cdf | cumulative distribution function |
| df or df's | distribution function (s) |
| $F$ | distribution function |
| $X_{k,n}$ | distribution function of the $k$th order statistics |
| $D\left(H_\gamma\right)$ | domain of attraction of $H_\gamma$ |
| $F_n$ | empirical distribution function |
| $\hat{U}_n$ | empirical tail function |
| $\overset{d}{=}$ | equality in distribution |

| | |
|---|---|
| EVI | Extreme Value Index |
| EVT | Extreme Value Theory |
| $E(X)$ | expectation of $X$ |
| ES | expected shortfall |
| e.g. | for example |
| $\Phi_\alpha$ | Fréchet distribution |
| FA | frequency analysis |
| GEV | Generalized Extreme Value |
| $F^{\leftarrow}$ | generalized inverse of $F$ |
| GPD | Generalized Pareto distribution |
| $\Lambda$ | Gumbel distribution |
| iff | if and only if |
| iid | independent and identically distributed |
| $\mathbb{I}_A$ | indicator function of a set $A$ |
| $\inf A$ | infimum of a set $A$ |
| i.e. | in other words |
| $n$ | integer greater than 1 |
| $f_{X_{j,n},X_{k,n}}$ | joint density of two order statistics $X_{j,n}$ and $X_{k,n}$ |
| KS | Kolmogorov-Smirnov test |
| $l_{\gamma,\sigma}(x_1,\cdots,x_{N_u})$ | log-likelihood function |
| ML | Maximum Likelihood |
| $L(\theta;X_1,\cdots,X_n)$ | Maximum likelihood function |
| $X_{n,n}$ | maximum of $X_1,X_2,\ldots,X_n$ |
| $\mu$ | mean of a rv |
| MM | method of moments |
| $X_{1,n}$ | minimum of $X_1,X_2,\cdots,X_n$ |
| $\mathbb{N}$ | natural numbers |
| $x$ | observation from $X$ |
| $U_{1,n},U_{2,n},\cdots,U_{n,n}$ | order statistics corresponding to $U_1,U_2,\cdots,U_n$ |
| $V_{i,n}$ | order statistics corresponding to a sample of rv's of standard Pareto distribution |
| $E_{1,n},\cdots,E_{n,n}$ | order statistics corresponding to a sequence of $n$ iid rv's |
| $(X_{1,n},\cdots,X_{n,n})$ | order statistics of $n$ iid observations from a rv $X$ |
| $\theta:=(\gamma,\mu,\sigma)$ | parameters of the GEV distribution |
| $\ominus$ | parameter space of the GEV distribution |

| | |
|---|---|
| POT | Peaks Over Threshold |
| $f_{X_{k,n}}$ | probability distribution function of $X_{k,n}$ |
| $f$ | probability density function |
| $h_\gamma$ | probability density function of the GEV |
| $h_\theta\left(x\right)$ | probability density function of the GEV with parameter $\theta$ |
| $M_{p,r,s}$ | Probability Weighted Moment |
| pdf | probability distribution function |
| $(\Omega, \mathcal{F}, P)$ | probability space |
| PWM | Probability Weighted Moment |
| $Q$ | quantile function |
| $x_p$ | quantile of order $p$ |
| rv or rv's | random variable (s) |
| $\mathbb{R}$ | real number |
| $h\left(t\right)$ | regularly varying functions |
| $\mathcal{R}_\rho$ | regularly varying functions |
| $R_m$ | return level |
| $T$ | return period |
| $\rho$ | risk measure |
| $X$ | rv |
| $U_1, U_2, \cdots, U_n$ | rv's from a uniform distribution on $[0, 1]$ |
| $V$ | rv under Fréchet distribution |
| $G$ | rv under Gumbel distribution |
| $W$ | rv under Weibull distribution |
| $V_1, V_2, \ldots, V_n$ | sequence of rv's under standard Pareto distribution |
| $A_n$ | simplex |
| $\sup A$ | supremum of a set $A$ |
| $E_1, E_2, \cdots, E_{n+1}$ | standard exponential distribution |
| $(X_1, X_2, \ldots, X_n)$ | sequence of iid rv's |
| SD | standard deviation |
| $H_\gamma$ | standard generalized extreme value distribution |
| s.t. | such that |
| SW | Shapiro-Wilk |
| $\mathcal{N}\left(0, 1\right)$ | standard Gaussian distribution |
| $L$ | slowly varying function |
| $u$ | threshold |

| | |
|---|---|
| TC | threshold choice |
| $\bar{F}$ | tail distribution |
| $V_n$ | uniform quantile function |
| $S_{i,n}$ | uniform spacing |
| $x_F$ | upper endpoint |
| $G_n$ | uniform empirical distribution |
| VaR | Value-at-Risk |
| $\Psi_\alpha$ | Weibull distribution |

# Bibliography

[1] Akaike, H. (1974). A new look at the statistical model identification. IEEE Trans. Automatic Control **19** (6) : 716–723.

[2] Arnold, B.C., Balakrishan, N. and Nagaraja, H.N. (1992). A First Course in Order Statistics. Wiley, New York.

[3] Artzner, P., Delbaen, F., Eber, J.M. and Heath, D. (1997). Thinking Coherently. Risk, **10** (11) : 68–71.

[4] Artzner, P., Delbaen, F., Eber, J.M. and Heath, D. (1999). Coherent measures of risk. Mathematical Finance, 203-28.

[5] Ashkar, F., Ouarda, T.B.M.J. (1993). Robust estimators in hydrologic frequency analysis. In: Kuo CY (ed) Engineering hydrology. Am. Soc. Civ. Eng., New York, 347–352.

[6] Balkema, A.A., de Haan, L. (1974). Residual lifetime at great age. Ann Probab **2** : 792–804

[7] Bayazıt, M., Shaban, F., Onoz, B. (1997). Generalized extreme value distribution for flood discharges. Turk J Eng Environ Sci **21** (2) : 69–73.

[8] Beirlant, J., Goegebeur, Y., Segers, J., Teugels, J., de Waal, D. and Ferro, C. (2004). Statistics of extremes: theory and applications. Wiley.

[9] Billingsley, P. (1995). Probability and Measure, 3rd edition. Wiley, New York.

[10] Benameur, S., Meraghni, D., Benkhaled, A., Chebana, F. and Necir, A. (2016). Complete Flood Frequency Analysis in Abiod Watershed Biskra (Algeria). Natural Hazards, Springer. vol 86, issue 2, page 519-534.

[11] Benazzouz, MT. (2010). Flash floods in Algeria: impact and management. G-WADI meeting. Flash Flood Risk Management Expert Group. Meeting Cairo, 25–27 September 2010, Egypt.

[12] Benkhaled, A., Higgins, H., Chebana, F. and Necir, A. (2014). Frequency analysis of annual maximum suspended sediment concentrations in Abiod wadi, Biskra (Algeria). Hydrol Process **28** (12) : 3841–3854.

[13] Beran, M.A., Nozdryn-Plotnicki, M.J. (1977). Estimation of low return period floods. Hydrol Sci Bull **22** (2) : 275–282.

[14] Bingham, N.H., Goldie, C.M. and Teugels, J.L. (1987). Regular Variation. Cambridge University Press, Cambridge.

[15] Bobée, B. (1999). Extreme flood events valuation using frequency analysis. A Crit Rev. Houille Blanche **54** : 100–105.

[16] Bobée, B. and Ashkar, F. (1991). The gamma family and derived distributions applied in hydrology. Water Resources Publications.

[17] Bouanani, A. (2005). Hydrologie, Transport solide et Modélisation. Etude de quelques sous bassins de la Tafna. Université de Tlemcen.

[18] Bühlmann, H. (1970). Mathematical Models in Risk Theory. Springer, New York.

[19] Bühlmann, H. (1980). An Economic Premium Principle. ASTIN Bulletin 11, 52-60.

[20] Brahimi, B. (2011). Statistics of bivariate extreme values. Doctorat thesis, Med Khider University, Biskra-Algeria.

[21] Castillo, E., Hadi, A.S., Balakrishnan, N. and Sarabia, J.M. (2005). Extreme Value and Related Models with Applications in Engineering and Science. Wiley series in probability and statistics.

[22] Chebana, F. and Ouarda, T.B.M.J. (2011). Multivariate quantiles in hydrological frequency analysis. Environmetrics **22** (1) : 63–78.

[23] Chebana, F., Ouarda, T.B.M.J., Bruneau, P., Barbet, M., El Adlouni, S. and Latraverse, M. (2009). Multivariate homogeneity testing in a northern case study in the province of Quebec Canada. Hydrol Process **23** : 1690–1700.

[24] Chebana, F., Adlouni, S. and Bobée, B. (2010). Mixed estimation methods for Halphen distributions with applications in extreme hydrologic events. Stoch Env Res Risk Assess **24** (3) : 359–376.

[25] Cheng, S. and Peng, L. (2001). Confidence Intervals for the Tail Index. Bernoulli **7**, 751-760.

[26] Chow, V.T., Maidment, DR., Mays, LW. (1988). Applied hydrology. McGraw-Hill, New York.

[27] Clarke, R.T. (1994). Fitting distributions. Chapter 4 Statistical modeling in hydrology. Wiley, 39–85

[28] Coles, S.G. (2001). An Introduction to Statistical Modelling of Extreme Values, Springer Series in Statistics.

[29] Csörgo, S., Deheuvels, P. and Mason, B. (1985). Kernel Estimator of the Tail Index of a Distribution. Annals of Statistics **13**, 1050-1077.

[30] Csörgo, S., and Mason, D.M. (1985). Central Limit Theorems for Sums of Extreme Values. Mathematical Proceedings of the Cambridge Philosophical Society 98, 547-558.

[31] Danielsson, J., de Haan, L., Peng, L. and de Vries, C.G. (2001). Using a Bootstrap Method to Choose the Sample Fraction in Tail Index Estimation. Journal of Multivariate Analysis, **76**, 226-248.

[32] David, H. A. and Nagaraja, H. N. (2003). Order Statistics, Third Edition. Wiely.

[33] Davis, R. and Resnick, S.I. (1984). Tail Estimates Motivated by Extreme Value Theory. Annals of Statistics 12, 1467-1487.

[34] Davison, A.C. and Smith, R.L. (1990). Models for exceedances over high thresholds, Journal of the Royal Statistical Society. Series B Methodological, **52**, N°3.

[35] Deheuvels, P., Häusler, E. and Mason, D.M. (1988). Almost Sure Convergence of the Hill Estimator. Mathematical Proceedings of the Cambridge Philosophical Society 104, 371-381.

[36] Dekkers, A.L.M., Einmahl, J.H.J. and de Haan, L. (1989). A Moment Estimator for the Index of an Extreme Value Distribution. Annals of Statistics **17**, 1833-1855.

[37] Dekkers, A.L.M. and de Haan, L. (1989). On the Estimation of the Extreme Value Index and Large Quantile Estimation. Annals of Statistics **17**, 1795-1832.

[38] Denuit, M., Dhaene, J., Goovaerts, M. and Kaas, R. (2005). Actuarial theory for dependent risks measures, orders and models. Wiley.

[39] Devroye, L. (1986). Non-Uniform Random Variate Generation. Springer-Verlag, New York, Berlin, Heidelberg Tokyo.

[40] Dubief, J. (1953). Essai sur l'hydrologie superficielle au Sahara. GGA, Direction du Service de la Colonisation et de l'Hydraulique, Service des Etudes Scientifiques, Alger, p 457.

[41] Draisma, G., de Haan, L., Peng, L. and Perreira, T.T. (1999). A Bootstrap-Based Method to Achieve Optimality in Estimating the Extereme-Value Index. Extremes **2**, 367-404.

[42] Drees, H. and Kaufmann, E. (1998). Selection of the Optimal Sample Fraction in Univariate Extreme Value Estimation. Stochastic Processes and their Applications **75**, 149-195.

[43] Ehsanzadeh, E., El Adlouni, S. and Bobee, B. (2010). Frequency analysis incorporating a decision support system for hydroclimatic variables. J Hydrol Eng **15** (11) : 869–881.

[44] El Adlouni, S. and Bobée, B. (2010). Système d'aide à la decision pour l'estimation du risqué hydrologique. IASH Publ, p 340.

[45] El Adlouni, S., Bobée, B. and Ouarda, T.B.M.J. (2008). On the tails of extreme event distributions in hydrology. J Hydrol 355:16–33.

[46] Ellouze, M. and Abida, H. (2008). Regional flood frequency analysis in Tunisia: identification of regional distributions. Water Resour Manag **22** (8) : 943–957.

[47] Embrechts, P., Klüppelberg, C. and Mikosch, T. (1997). Modelling extremal events for insurance and finance. Springer, Heidelberg.

[48] Fisher, R.A. and Tippett, L.H.C. (1928). Limiting Forms of the Frequency Distribution of the Largest or Smallest Member of a Sample. Proceedings of the Cambridge Philosophical Society **24**, 180-190.

[49] Geluk, J.L. and de Haan, L. (1987). Regular Variation, Extensions and Tauberian Theorems. Mathematical Centre Tract 40, Centre for Mathematics and Computer Science, Amsterdam.

[50] Gnedenko, B.V. (1943). Sur la Distribution Limite du Terme Maximum d'une Série Aléatoire. Annales de Mathématiques **44**, 423-453.

[51] Greenwood, J.A. Landwehr, J.M. Matalas, N.C. and Wallis, J.R. (1979). Probability weighted moments: definition and relation to parameters of several distributions expressable in inverse form. Water Resources Research **15**, 1049–1054.

[52] Groeneboom, P., Lopuhaä, H.P.and de Wolf, P.P. (2003). Kernel Estimators for the Extreme Value Index. Annals of Statistics **31**, 1956-1995.

[53] Grubbs, FE. and Beck, G. (1972). Extension of sample sizes and percentage points for significance tests of outlying observations. Technometrics **14** : 847–854.

[54] Gumbel, E.J. (1958). Statistics of Extremes. Columbia University Press, New York.

[55] de Haan, L. and Stadtmüller, U. (1996). Generalized Regular Variation of Second Order. Journal of the Australian Mathematical Society (Series A) 61, 381-395.

[56] de Hann, L. and Rootzen, H. (1993). On the Estimation of High Quantiles. Journal os statistical planning and Inference **35**, 1-13.

[57] de Haan, L. and Ferreira, A. (2006). Extreme value theory: an introduction. Springer Series in Operations Research and Financial Engineering, Boston.

[58] de Haan, L. and Peng, L. (1998). Comparison of Tail Index Estimators. Statistica Neerlandica 52, 60-70.

[59] Haktanir, T. (1992). Comparison of various flood frequency distributions using annual flood peaks data of rivers in Anatolia. J Hydrol **136** : 1–31.

[60] Hall, P. (1990). Using the Bootstrap to Estimate Mean Squared Error and Select Smoothing Parameter in Nonparametric Problems. Journal of Multivariate Analysis **32**, 177-203.

[61] Hall, P. (1982). One Some Simple Estimates of an Exponent of Regular Variation. Journal of the Royal Statisticol Society **44**, 37-42.

[62] Hall, P. and Welsh, A.H. (1985). Adaptive Estimates of Parameters of Regular Variation. Annals of Statistics **13**, 331-341.

[63] Hamel, A. (2009). Hydrogéologie des systèmes aquifères en pays montagneux à climat semi-aride. Cas de la vallée d'Oued El Abiod (Aurès). Université, Mentouri, Constantine.

[64] Hebal, A. and Remini, B. (2011). Choix du modèle fréquentiel le plus adéquat à l'estimation des valeurs extrêmes de crues (cas du nord de L'Algérie). Can J Civ Eng **38** (8) : 881–892.

[65] Hill, B. (1975). A Simple General Approach to Inference About the Tail of a Distribution. Annals of Statistics **3**, 1163-1174.

[66] Hosking, J.R.M. (1990). L-moments: analysis and estimation of distributions using linear combinations of order statistics. J R Stat Soc **52** : 105–124.

[67] Hosking, J.R.M. and Wallis, J.R. (1987). Parameter and quantile estimation for the generalized Pareto distribution. Technometrics **29** : 339–349.

[68] Hosking, J.R.M. and Wallis, J.R. (1997). Regional frequency analysis; an approach based on L-moments. Cambridge University Press, Cambridge, p 224.

[69] Hosking, J.R.M. and Wallis, J.R. and Wood, E.F. (1985). Estimation of the generalized extreme value distribution by the method of probability-weighted moments. Technometrics **27** : 257–261.

[70] (IACWD), I. C. o. W. D. (1982). Guidelines for determining flood flow frequency: bulletin 17B (revised and corrected). Hydrol. Subcomm, Washington, DC, p 28.

[71] Ihaka, R., Gentleman, R. (1996). R: a language for data analysis and graphics. J Comput Graph Stat **5** (3) : 299–314.

[72] Institute of Hydrology (IH). (1999). Flood estimation handbook. Wallingford, UK, p 1999.

[73] Jenkinson, A.F. (1955). The Frequency Distribution of the Annual Maximum (or Minimum) of Meteorological Elements. Quarterly Journal ofthe Royal Meteorological Society **81**,158-171.

[74] Jones, B. L. and Zitikis, R., 2003. Empirical estimation of risk measures and related quantities. North American Actuarial Journal. 7(4), 44-54.

[75] Kaas, R. Goovaerts, M. Dhaene, J. and Denuit, M. (2009).Modern Actuarial Risk Theory Using R. Springer.

[76] Kaiser, T., Brazauskas, V., (2006). Interval estimation of actuarial risk measures. North American Actuarial Journal, Volume 10, Number 4.

[77] Karim, M.A. and Chowdhury, J.U. (1995). A comparison of four distributions used in flood frequency analysis in Bangladesh. Hydrol Sci J **40** (1) : 55–66.

[78] Khamis, H.J. (1997). The delta-corrected Kolmogorov–Smirnov test for the two-parameter Weibull distribution. J Appl Stat **24** (3) : 301–318.

[79] Kite, G.W. (1988). Flood and risk analyses in hydrology. Water Resources Publications, Littleton, p 187.

[80] Kotz, S. and Nadarajah, S. (2000). Extreme Value Distributions: Theory and Applications. London: Imperial College Press.

[81] Koutsoyiannis, D. (2003). On the appropriateness of the Gumbel distribution for modeling extreme rainfall. In: Brath A, Montanari A Toth E (eds) Hydrological risk: recent advances in peak river flow modelling, prediction and realtime forecasting. Assessment of the impacts of land-use and climate changes. Editoriale Bios, Castrolibero, Bologna, 303–319.

[82] Laio, F. (2004). Cramer–von Mises and Anderson–Darling goodness of fit tests for extreme value distribution with unknown parameters. Water Resour Res W09308:40.

[83] Landwehr, J.M., Matalas, N.C. and Wallis, J.R. (1979). Probability weighted moments compared with some traditional techniques in estimating Gumbel parameters and quantiles. Water Resources Research **15**, 1055–1064.

[84] Lang, M., Ouarda, T.B.M.J. and Bobée, B. (1999). Towards operational guidelines for over-threshold modeling. J Hydrol **225** : 103–117.

[85] Liao, M. and Shimokawa, T. (1999). A new goodness-of-fit test for type-I extreme-value and 2-parameter Weibull distributions with estimated parameters. J Stat Comput Simul **64** (1) : 23–48.

[86] Mason, D.M. (1982). Laws of Large Numbers for Sums of Extreme Values. Annals of Probability 10, 756-764.

[87] Malmquist, S. (1950). On a proprety of order statistics from a rectangular distribution. Skandinavish Aktuarietidkrift **33**, 214-222.

[88] Mebarki, A. (2005). Hydrologie des bassins de l'Est Algérien. Thèse de doctorat, université de Constantine, Constantine.

[89] Meraghni, D. (2008). Modeling Distribution Tails. Doctorat thesis, Med Khider University, Biskra-Algeria.

[90] Mutua, F.M. (1994). The use of the Akaike information criterion in the identification of an optimum flood frequency model. Hydrol Sci J **39** (3) : 235–244.

[91] National Research Council (NRC). (1988). Estimating probabilities of extreme floods: methods and recommended research. National Academy Press.

[92] Natural Environment Research Council (NERC). (1975). Flood studies report. London, p 1.

[93] Natural Environment Research Council (NERC). (1999). Flood studies report, in five volumes. Institute of Hydrology, Wallingford.

[94] Necir, A. (2006). A Functional Law of the Iterated Logarithm for Kernel-type Estimators of the Tail Index. Journal of Statistical Planning and Inference **136**, 780-802.

[95] Neves, C. and Fraga Alves, M.I. (2004). Reiss and Thomas' Automatic Selection of the Number of Extremes. Computational Statistics and Data Analysis **47**, 689-704.

[96] Ouarda, T.B.M.J., Ashkar, F., Bensaid, E. and Hourani, I. (1994). Statistical distributions used in hydrology. Transformations and asymptotic properties. University of Moncton, Moncton.

[97] Pickands, J. (1975). Statistical Inference Using Extreme Order Statistics. Annals of Statistics **3**, 119-131.

[98] Pyke, R. (1965). Spacings. J. Roy. Statis. Soc. **B 27**, 395-449.

[99] Pyke, R. (1972). Spacings Revisted Proceedings of the Sixth Berkeley Symposium on Mathematical statistics and probability, vol **1**, 414-427. University of California Press, Berkeley.

[100] Rao, A.R., Hamed, KH. (2000). Flood Frequency Analysis. City

[101] Rasmussen, P.F., Ashkar, F., Rosbjerg, D. and Bobée, B. (1994). The POT method for flood estimation: a review. In: Hipel KW (ed) Stochastic and statistical methods in hydrology and environmental engineering, extreme values: floods and droughts. Kluwer Academic, Dordrecht. Water Sci Technol Libr **1** : 15–26.

[102] Reiss, R.D. (1989). Approximate distributions of order statistics. Springer, New York.

[103] Reiss, R.D., Thomas, M. (1997). Statistical analysis of extreme values: with applications to insurance, finance, hydrology and other fields. Birkhäuser Verlag, Basel-Boston-Berlin

[104] Resnick, S.I. (2007). Heavy-Tail Phenomena, probabilistic and statistical modeling. Springer.

[105] Salas, J.D., Smith, R. (1980). Computer programs of distribution functions in hydrology. Colorado State University, Fort Collins.

[106] Saporta, G. (1990). Probabilités, Analyse des Données et Statistique. Editions Technip, Paris.

[107] Scholz, F.W., Stephens, M.A. (1987). K-sample Anderson–Darling tests. J Am Stat As **82** (399) : 918–924.

[108] Schwartz, G. (1978). Estimating the dimension of a model. Ann Stat **6** : 461–464.

[109] Smith, R.L. (1985). Maximum Likelihood Estimation in a Class of Non Regular Cases. Biometrika **72**, 67-90.

[110] Smith, R.L. (1987). Estimating Tails of Probability Distributions. Annals of statistics **15**, 1174-1207.

[111] Strassen,V. (1964). An invariant principal for the law of the iterated logarithm. Z.Wahrscheinlichkeitstheorie. verw. Gebiete. **3**, 211–226.

[112] Sukatme, P.V. (1937). Tests of significance for sample of the chi square population with two degrees of freedom. Ann, Eugen **8**, 52-56.

[113] Tanaka, S and Takara, K. (2002). A study on threshold selection in POT analysis of extreme floods. IAHS-AISH publication, p 271.

[114] Todorovic, P., Zelenhasic, E. (1970). A stochastic model for flood analysis. Water Resour Res **6** (6) : 1641–1648.

[115] Tramblay, Y., St-Hilaire, A., Ouarda, T.B.M.J. (2008). Frequency analysis of maximum annual suspended sediment concentrations in North America. Hydrol Sci J **53** (1) : 236–252.

[116] Vogel R.M, McMahon TA, Chiew FHS (1993). Floodflow frequency model selection in Australia. J Hydrol **146** : 421–449.

[117] von Mises, R. (1936). La Distribution de la plus grande des n valeurs. Selected papers, American Mathematical Society, 271-294.

[118] Wang, S.S. (1996). Premium Calculation by Transforming the Layer Premium Density. ASTIN Bulletin 26, 71-92.

[119] Weissman, I. (1978). Estimation of parameters and large quantiles based on the k largest observations. J. Amer. Statist. Assoc., **73** ,812-815.

[120] Willeke, G.E., Hosking, J.R., Wallis, J.R. and Guttman, N.B. (1995). The national drought atlas (draft), IWR Rep US Army, Corps of Eng, 94-NDS-4.

[121] Yaari, M.E. (1987). The Dual Theory of Choice under Risk. Econometrica 55, 95-115.

[122] Young, V. (2004). Premium Principles. Encyclopedia of Actuarial Science. Wiley 4, 1322-1331.